

Information-Theoretic Viewpoints on Optimal Causal Coding-Decoding Problems

Siva K. Gorantla and Todd P. Coleman
ECE Department
Coordinated Science Laboratory
University of Illinois
Urbana, IL
{sgorant2,colemant}@illinois.edu

Abstract—In this paper we consider an interacting two-agent sequential decision-making problem consisting of a Markov source process, a causal encoder with feedback, and a causal decoder. Motivated by a desire to foster links between control and information theory, we augment the standard formulation by considering general alphabets and a cost function operating on current and previous symbols. Using dynamic programming, we provide a structural result whereby an optimal scheme exists that operates on appropriate sufficient statistics. We emphasize an example where the decoder alphabet lies in a space of beliefs on the source alphabet, and the additive cost function is a log likelihood ratio pertaining to sequential information gain. We also consider the inverse optimal control problem, where a fixed encoder/decoder pair satisfying statistical conditions is shown to be optimal for some cost function, using probabilistic matching. We provide examples of the applicability of this framework to communication with feedback, hidden Markov models and the nonlinear filter, decentralized control, brain-machine interfaces, and queuing theory.

I. Introduction

Many current and future societal problems involve designing and understanding networks of sequential decision-making entities cooperating in an uncertain environment. Some of these entities may be physical/biological agents, whereas others might be computerized systems. For example, cyber-physical systems feature interacting networks of physical processes that are noisily sensed and actuated by computational algorithms. The mammalian brain comprises another example, where the cooperative goals of sensing, perception, learning, and eliciting behavior are achieved via coupled neural systems that interact via signaling across a noisy biological medium.

From an engineering system designer vantage point, obtaining optimal coordination strategies for a network of interacting decision-makers is in general computationally intractable [1]. For a class of small networks (e.g. comprising a specific interaction structure between an encoder and a decoder), and an asymptotic performance objective, fundamental limits of performance can be addressed using the information theoretic concepts of communication and rate-distortion [2]. Identifying optimal strategies for sequential decision-making under uncertainty for a single agent, on the flipside, is traditionally addressed with control theoretic-principles of Markov decision theory [3].

From a scientific vantage point, the joint statistical dynamics between interacting decision-makers can provide insight into the cost or utility they are cooperatively optimizing. For small networks (e.g. an encoder and decoder) with a limited statistical dynamics interaction structure, this has been addressed with the information-theoretic principle of source-channel probabilistic matching [4]. Inverse optimal control theory theory identifies cost functions for which a fixed strategy of one decision-maker is optimal [5] and has been used in neural [6], [7] and cognitive science [8] applications.

It appears evident that understanding this class of problems for more general objectives and interaction structures can utilize insights from both information and control theory, but the differences in their philosophical starting points is striking: **Information Theory** problems, traditionally specify large but *fixed* time horizon n for which some decisions are not made until this terminal point. Even in problems where neither an observation nor a decision variable lies in a time-horizon dependent set (e.g. reproducing a source over a noisy channel with a fidelity criterion), Shannon's 'separation theorem' [9] shows that for very large n , it is sufficient to first decompose the problem into sub-problems, each of which contains some observations or decision variables with time horizon-dependent alphabet structure (e.g. of size 2^{nR}) and a performance objective pertaining to constrained extremizing of mutual information. As such, traditional information theoretic problem formulations have the following starting point:

- (a) *time horizon-dependent alphabets*
- (b) *some decisions made at final stage of long time horizons*
- (c) *performance objective: extremize mutual information*

Control Theory Markov decision theory problems typically involve observations of state variables' whose future statistics are impacted by their current values and the current 'decision variable' that is under causal control of a decision-maker. The alphabet size of observations and decision variables are typically *unrelated* to the time horizon n of the problem. Moreover, at each time step, a decision must be made based upon causal information up to that time. Lastly, the performance objective is to minimize an expected sum of costs, each of which operates on current state, observation, and decision variables. Structural results are typically desirable in such settings because they develop conditions relating the existence

of explicit, non-random strategies that operate on sufficient statistics. Succinctly, we can state this as follows:

- (a) *time horizon-independent alphabets*
- (b) *decisions made sequentially based on causal information*
- (c) *performance objective: sum of costs operating on current observations and decision variables*

So these two philosophies have striking differences. Consider the class of ‘causal coding/decoding’ problem that further demonstrates this: At each time step i , the causal encoder’s

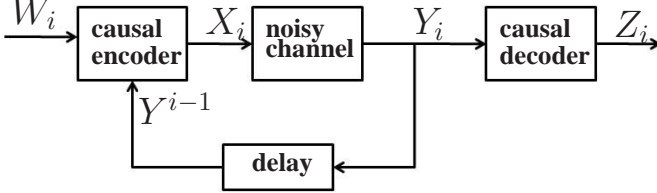


Fig. 1. Basic problem setup: an optimal causal coding/decoding problem.

decision variable is the input $X_i \in \mathcal{X}$ to a noisy channel that is a causal function of source inputs (W_1, \dots, W_i) and the noisy channel outputs Y_1, \dots, Y_{i-1} : $X_i = e_i(W^i, Y^{i-1})$. The causal decoder’s decision variable is a ‘source estimate’ $Z_i \in \mathcal{Z}$ that is a causal function of channel outputs (Y_1, \dots, Y_i) : $Z_i = d_i(Y^i)$. They jointly design their strategies $\pi = (e, d)$ to minimize a function $J_{n,\pi}$ pertaining to an expected sum of costs:

$$J_{n,\pi} = \mathbb{E}_{e,d} \left[\sum_{i=1}^n g(W_i, Z_i) \right] \quad (1)$$

Some aspects of the problem appear to make it amenable to a control theoretic analysis: (a) the source alphabet \mathcal{W} is unrelated to the time horizon n and (b) the sequential decision-making and additive costs, (c) the performance objective (1) operates additively on observations/decision variables in the vicinity of each time i as compared to only at the final time horizon n . The presence of the noisy channel in the loop possibly make it amenable to an information-theoretic analysis: mutual information could plausibly provide tight bounds on attainable costs. On the flipside, neither agent’s observations at any time point are a nested version of the other’s and so they have a ‘non-classical’ information structure [10] - making this a ‘hard’ control problem. Analogously, the ‘hard’ delay constraint pertaining to causal decoding and typical ‘hard decision’ assumption of \mathcal{W}, \mathcal{Z} being in discrete, time-horizon independent alphabets typically render information-theoretic techniques irrelevant to the understanding of these ‘real-time’ problems [11], [12].

In this paper we consider a causal coding/decoding problem where W is a Markov source process. We consider additive cost functions operating of the form $g(w_i, x_i, z_{i-1}, z_i)$. We do not impose assumptions (e.g. finiteness) on alphabets of the variables. Our motivation for this more general framework is an example (Section I-A) motivated by feedback communication where the source alphabet is continuous, the decoder alphabet lies in a space of *beliefs* on the source alphabet, and the additive cost function is a log likelihood ratio pertaining to *sequential information gain*. Using dynamic programming,

we provide a structural result whereby an optimal scheme exists that operates on appropriate sufficient statistics. We also consider the inverse optimal control problem, where a fixed encoder/decoder pair satisfying a sufficient statistical condition is shown to be optimal for some cost function, using probabilistic matching. We provide examples of the applicability of this framework to communication with feedback, hidden Markov models and the nonlinear filter, decentralized control, brain-machine interfaces, and queuing theory.

A. Example: Communication over a Noisy Channel with Feedback and the Sequential Information Gain Cost

We now consider the traditional feedback communication model and how its assumptions - along with traditional ‘real-time’ problem assumptions - can be modified so that fundamental limits are unchanged but the frameworks align. Consider the traditional information-theoretic communication model with feedback, consisting of an encoder, a decoder, and a fixed block length n . The encoder has a message $W \in \mathcal{W} = \{1, \dots, 2^{nR}\}$. It specifies n inputs to the channel, X_1, \dots, X_n . The channel is memoryless and non-anticipative where $P_{Y|X}(y|x)$ is the statistics of the output given the input. At each time step i , the encoder selects the message W and the previous channel outputs Y_1, \dots, Y_{i-1} at time i , to specify the next channel input X_i . The decoder, at time n , having acquired channel outputs Y_1, \dots, Y_n , specifies a single decision, $\hat{W}_n \in \mathcal{W}$. The question asked in information theory is, how large can R be such that for sufficiently large n , there exist encoders and decoders for which $\mathbb{P}(\hat{W}_n \neq W) \rightarrow 0$? To demonstrate the existence of such encoders and decoders, a *random coding* argument [2] and the laws of large numbers are typically invoked.

Recently, a development by Shayevitz & Feder [13], [14], [15], has re-visited a philosophically different way to frame the feedback communication model - dating back to the 1960s [16], [17], [18] - that has a more dynamical systems and control theoretic flavor. Consider the following changes to the standard

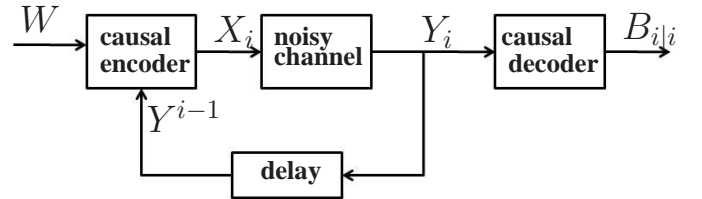


Fig. 2. Communication of a message point W with causal feedback over a memoryless channel.

information theoretic formulation that more closely resembles a causal coding/decoding problem, shown in Figure 2:

- **Message Point:** $W_i = W$ is equally likely over interval $\mathcal{W} = [0, 1]$.
- **Decoder:** At each i (not only at time n), the decoder specifies $Z_i = B_{i|i}$, the posterior belief about W given Y_1, \dots, Y_i : $B_{i|i}(A) \triangleq \mathbb{P}(W_i \in A | Y^i)$.
- **Achievability:** As shown in Fig 3, with a set of uniform quantizers $(q_{iR} : [0, 1] \rightarrow 2^{iR}, i \geq 1)$, a rate R is achievable if $B_{i|i}(\{w : q_{iR}(w) = q_{iR}(W)\}) \rightarrow 1$.

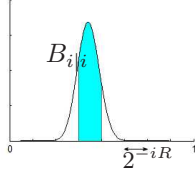


Fig. 3. Representation of the posterior belief B_i in terms of its density.

Note the importance of W being a *continuous interval* and Z being the space of beliefs on W , $\mathcal{P}(W)$, in order for this ‘real-time’ flavored problem to relate to traditional information-theoretic notions of achievability. The fundamental limits under both formulations are equivalent [15], where achieving capacity subject to channel input cost $\eta(x)$ constraints pertains to maximizing the mutual information $I(W; Y^n)$ [9]. A time-invariant ‘posterior matching’ encoding scheme in Figure 2’s framework achieves capacity on general memoryless channels [15]. Moreover, it is an optimal solution to a stochastic control problem [19] whose cost function at each time step is related to the *sequential information gain* $I(W; Y_i|Y^{i-1})$:

$$I(W; Y^n) = \sum_{i=1}^n I(W; Y_i|Y^{i-1}) = \sum_{i=1}^n \mathbb{E} \left[\log \frac{dB_{i|i}}{dB_{i-1|i-1}}(W) \right]$$

Note that the sequential information gain term represents the reduction in W ’s uncertainty from the previous posterior belief $B_{i-1|i-1}$ to the current, and so each term in the sum operates on W , $B_{i|i}$, and $B_{i-1|i-1}$. This alludes to a generalization of causal coding/decoding problems with a cost function $g(w_i, x_i, z_{i-1}, z_i)$, which in this case could plausibly be

$$g(w_i, x_i, z_{i-1}, z_i) = -\log \frac{dz_i}{dz_{i-1}}(w_i) + \alpha\eta(x_i), \quad (2)$$

where $Z_i \in Z = \mathcal{P}(W)$ is a *decision variable* that can be any belief about the message. In this manuscript, we plan to build on this example and formulate general problems that capture this generalization and further elucidate an interplay between information theory and control theory within the context of both designing optimal strategies and performing inverse optimal control to characterize cost functions for which fixed strategies are optimal.

B. Related Work

The interplay between information and control theory has been established when treating a ‘message point’ as a real-valued point within the context of control over noisy channels [20], [21], [22], [23], [24], [25], [26] and feedback information theory [16], [17], [18], [13], [15], [27].

Cost functions pertaining to log likelihood ratios and making decision variables pertaining to beliefs have been used within the context of sequential prediction [28], [29] and signal compression/classification [30], relating thermodynamics and information theory with inference on hidden Markov models [31], [32], linearly solvable Markov decision problems [33], and feedback information theory [34], [19], [35], [36].

Causal coding-decoding problems akin to Figure 1 have been studied extensively when W is a Markov process, as it

typically enables the use of dynamic programming to demonstrate the existence of optimal strategies where agents use posterior beliefs as state variables. In the case of all-discrete alphabets, this was demonstrated in [37],[38]. In the case of all-real alphabets, W a Gauss-Markov process, and $P_{Y|X}$ an additive Gaussian channel, this was demonstrated in [39][20, Ch. 6] where additionally an explicit optimal scheme consisting of ‘innovations-encoding’ and ‘minimum-mean-squared error decoding’ strategy was constructed. [36] considered the case of W discrete and the objective to maximize mutual information, but Z_i was not a decision variable. The case of general alphabets W was considered in [40], [41], but the purpose was quantization and thus Z was discrete and the cost function balanced squared error distortion and quantizer output entropy rate. Note that the information gain scenario in Section I-A does not fall within any of the aforementioned works.

Control-theoretic approaches to inverse optimal control involving a single-agent system have been developed classically for the case of a known policy [5], [42] where a control-Lyapunov function acts as an optimal value function and imposes constraints on candidate cost functions. Inverse reinforcement learning additionally requires inferring the single-agent’s policy based on experimental data [43], [44], [45], and it has been applied to solve extremely challenging engineering problems [46] and within the context of neural [6], [7] and cognitive [8] science.

Information-theoretic approaches to inverse optimal control for a two-agent (encoder/decoder) system relate costs to likelihood ratios through the variational equations for the rate-distortion and capacity-cost function [47], [48], but the problem formulations do not consider cost functions of the form $g(w_i, x_i, z_{i-1}, z_i)$ and either consider encoder/decoder interactions [49], [4] that do not have dynamics akin to the feedback loop and random process input in Figure 1 or have very specific statistical assumptions (e.g. the Gauss-Markov source and additive Gaussian channel [39][20, Ch. 6]).

C. Paper Outline and Main Results

We now outline the paper, where in each section we provide bullet points about how it differs from other formulations and its main results.

Section II provides mathematical notation and definitions that will be used throughout the manuscript.

Section III provides the problem setup. We emphasize the following properties that make it differ from traditional approaches:

- the Markov process source has a general alphabet W
- the traditional cost function $g(w_i, z_i)$ is replaced by

$$g(w_i, x_i, z_{i-1}, z_i) = \rho(w_i, z_{i-1}, z_i) + \alpha\eta(x_i) \quad (3)$$

- decision variables lie in arbitrary spaces X and Z

Section IV considers a fixed cost function (3) and finding optimal coordination strategies (e, d) . Results include:

- a structural result demonstrating the existence of optimal coordination strategies operating on sufficient statistics, capturing traditional results [37] as a special case.

Section V considers the *sequential information gain* cost function (2) with $Z = \mathcal{P}(W)$ and finding optimal coordination strategies. Results include:

- an optimal coordination strategy always specifies $Z_i = B_{i|i}$
- a characterization of the problem as cost-penalized maximization of mutual information $I(W^n; Y^n)$

The first result uses dynamic programming and the second law of thermodynamics for Markov chains [50]. It synergizes with work in [31] but differs in how this is cast in the causal coding/decoding framework and the information gain cost (2).

Section VI considers the inverse optimal control scenario with cost functions of the form (3). Results demonstrate that a ‘stationary Markov’ coordination strategy (\bar{e}, \bar{d}) is inverse optimal when the channel outputs Y_1, \dots, Y_n are statistically independent. The technique constructs the induced ρ and η from the variational equations to the rate-distortion and capacity-cost functions and can be interpreted as a ‘source-channel matching’ [49], [4] generalization applicable to the causal coding/decoding problem with cost function (3). It is also shown how in some situations, this sufficient condition reduces to time-reversibility of a Markov chain, thus further demonstrating a relationship between thermodynamics and information theory that has been developed in [51], [32].

Section VII, provides example problems for which the aforementioned results apply, and shows how:

- under a particular constraint, the hidden Markov model and nonlinear filter [52] are an optimal coordination strategy for the information gain cost (2) with $W = X$
- the posterior matching scheme [15] is an optimal coordination strategy for the information gain cost (2) and source model $W_i = W_{i-1}$ with $W = [0, 1]$
- the structural results aid the design of optimal and ‘user-friendly’ coordination strategies for brain-machine interfaces [53]
- inverse control optimal Markov coordination policies with cost (3) exist for:
 - Gauss-Markov source, AGN channel pair
 - Markov counting-function source, Z channel pair
 - Markov counting-function source, ‘inverted E ’ channel pair

The first example is related to the variational characterization of the optimality of the nonlinear filter [31], but is different due to the information gain cost (2). The second example generalizes the result of [19] because here, Z_i is a decision variable. In the inverse optimal control examples, the Gaussian channel case pertains to the decentralized control problems in [20, Ch 6], [39] with quadratic state cost and squared error distortion; the Z channel case pertains to the $\cdot/M/1$ queue for timing channels [54], [55]. the ‘inverted E ’ channel pertains to Blackwell’s trapdoor communication channel [56], [57], [58].

Section VIII provides a discussion and conclusion, followed by references and an appendix of proofs.

II. Definitions and Notations

Probabilistic Notation

- For a sequence a_1, a_2, \dots , denote a_i^j as (a_i, \dots, a_j) and $a^j \triangleq a_1^j$.
- Denote the probability space with sample space Ω , sigma-algebra \mathcal{F} , and probability measure \mathbb{P} as $(\Omega, \mathcal{F}, \mathbb{P})$.
- For a given (Ω, \mathcal{F}) and a Borel space $(V, \mathcal{B}(V))$, denote any measurable function $X : \Omega \rightarrow V$ as a random object. If $V = \mathbb{R}$, then X is termed a random variable.
- Upper-case letters V represent random objects and lowercase letters $v \in V$ represent their realizations.
- For two probability measures \mathbb{P} and \mathbb{Q} on (Ω, \mathcal{F}) , we say that \mathbb{P} is absolutely continuous with respect to \mathbb{Q} (denoted by $\mathbb{P} \ll \mathbb{Q}$) if $\mathbb{Q}(A) = 0$ implies $\mathbb{P}(A) = 0$ for all $A \in \mathcal{F}$. If $\mathbb{P} \ll \mathbb{Q}$, denote the Radon-Nikodym derivative as any random variable $\frac{d\mathbb{P}}{d\mathbb{Q}} : \Omega \rightarrow \mathbb{R}$ that satisfies

$$\mathbb{P}(A) = \int_{\omega \in A} \frac{d\mathbb{P}}{d\mathbb{Q}}(\omega) \mathbb{Q}(d\omega), \quad A \in \mathcal{F}.$$

- Denote $\mathcal{P}(V)$ as the space of probability measures on $(V, \mathcal{B}(V))$. For any random object $V : \Omega \rightarrow V$, denote $P_V(A) \triangleq \mathbb{P}(V \in A) \triangleq \mathbb{P}(\{\omega : V(\omega) \in A\})$, $A \in \mathcal{B}(V)$.
- Denote the conditional probability distribution of one random object V given that another U takes on u as

$$P_{V|U=u}(A) \triangleq \mathbb{P}(V \in A | U = u), \quad A \in \mathcal{B}(V).$$

Markov Chains Notation:

- A random process $V = (V_i : i \geq 1)$ is a Markov chain if

$$P_{V_{i+1}|V^i=v^i}(A) = P_{V_{i+1}|V_i=v_i}(A), \quad A \in \mathcal{B}(V). \quad (4)$$

It is *time-homogenous* if $P_{V_{i+1}|V_i=v_i}(A) = Q(A|v_i)$.

- A Markov chain is *time-reversible* if the forward and reverse time processes are statistically indistinguishable:

$$(V_j : 1 \leq j \leq n) \stackrel{d}{=} (V_{n-j+1} : 1 \leq j \leq n) \quad (5)$$

where $\stackrel{d}{=}$ denotes equivalence in distribution.

Information Theoretic Notation:

- Given two probability measures $P, Q \in \mathcal{P}(V)$, define the *Kullback-Leibler divergence* as

$$D(P||Q) \triangleq \begin{cases} \int_V \log \frac{dP}{dQ}(v) P_V(dv), & \text{if } P \ll Q \\ +\infty, & \text{otherwise} \end{cases} \quad (6)$$

- Given two sets of conditional distributions $(P_{V|U=u}, P'_{V|U=u} \in \mathcal{P}(V) : u \in \mathcal{U})$ and a distribution $P_U \in \mathcal{P}(U)$, define the *conditional divergence* as

$$D(P_{V|U} || P'_{V|U} | P_U) \triangleq \int_{\mathcal{U}} D(P_{V|U=u} || P'_{V|U=u}) P_U(du) \quad (7)$$

- Consider a set of conditional distributions $(P_{V|U=u} \in \mathcal{P}(V) : u \in \mathcal{U})$ and a distribution $P_U \in \mathcal{P}(U)$. This induces a marginal distribution $P_V \in \mathcal{P}(V)$. The mutual information is given by

$$I(P_{V|U}, P_U) \triangleq I(V; U) \triangleq D(P_{V|U} || P_V | P_U). \quad (8)$$

U and V are independent if and only if $I(V; U) = 0$.

- The conditional mutual information is given by

$$I(W; Y_2 | Y_1) = D(P_{W|Y_1, Y_2} || P_{W|Y_1} | P_{Y_1, Y_2}). \quad (9)$$

- The chain rule for mutual information is given by

$$I(W; Y^n) = \sum_{i=1}^n I(W; Y_i | Y^{i-1}). \quad (10)$$

$$\Rightarrow I(W^n; Y^n) = \sum_{i=1}^n I(W^n; Y_i | Y^{i-1})$$

- Consider a memoryless channel $P_{Y|X} = (Q_{Y|X=x} \in \mathcal{P}(Y) : x \in X)$, a cost function $\eta : X \rightarrow \mathbb{R}_+$, and an upper bound $L \in \mathbb{R}_+$. Define the capacity-cost function as $C(\eta, P_{Y|X}, L)$ [59] and its maximizing distribution $P_X^*(\eta, P_{Y|X}, L)$ as:

$$P_X^*(\eta, P_{Y|X}, L) \triangleq \arg \max_{P_X \in \mathcal{P}(X) \text{ s.t. } \mathbb{E}[\eta(X)] \leq L} I(P_X, P_{Y|X}) \quad (11)$$

$$C(\eta, P_{Y|X}, L) \triangleq I(P_X^*(\eta, P_{Y|X}, L), P_{Y|X}). \quad (12)$$

III. Problem Setup

Throughout this discussion, we consider 4 random processes W, X, Y, Z associated with Borel metric spaces W, X, Y, Z that are coupled according to Figure 1. The natural time ordering for the causal construction of the four random objects through time is given by:

$$\dots, Z_{i-1}, \underbrace{W_i, X_i, Y_i, Z_i}_{i\text{th epoch}}, W_{i+1}, \dots$$

The input process:

W is a time-homogenous Markov process such that for any $A \in \mathcal{F}_W$:

$$P_{W_{i+1}|W^i=w^i, X^i=x^i, Y^i=y^i}(A) = P_{W_{i+1}|W_i=w_i}(A) \quad (13)$$

$$\equiv Q_W(A|w_i) \quad (14)$$

The causal encoder:

The *causal encoder* at time i has causal information about the source, W^i , and causal feedback about the channel outputs, Y^{i-1} , to specify the next channel input, X_i ,

$$x_i = e_i(w^i, y^{i-1}). \quad (15)$$

We define the aspect of $e_i \in E_i$ that maps W_i to X_i as $\tilde{e}_i \in \tilde{E}$ where \tilde{E} is a space of Borel-measurable functions $f : W \rightarrow X$:

$$\tilde{e}_i(w^{i-1}, y^{i-1})(\cdot) = e_i\left(\left[\begin{smallmatrix} \cdot \\ w^{i-1} \end{smallmatrix}\right], y^{i-1}\right) \equiv \tilde{e}_i(\cdot). \quad (16)$$

and we define E_i to be the space of Borel-measurable functions $f : W^i \times Y^{i-1} \rightarrow X$ such that $\tilde{e}_i \in \tilde{E}$ for all w^{i-1} and y^{i-1} .

The memoryless non-anticipative channel:

$X_i \in X$ is passed through a time-homogenous, non-anticipative, memoryless channel to produce $Y_i \in Y$; for any $A \in \mathcal{F}_Y$:

$$P_{Y_i|Y^{i-1}=y^{i-1}, X^n=x^n, W^n=w^n}(A) = P_{Y|X}(A|x_i). \quad (17)$$

The causal decoder:

Lastly, the causal decoder at time i uses causal channel outputs, Y^i to specify $Z_i \in Z$. Define D_i as a space of Borel-measurable functions $f : Y^i \rightarrow Z$ and $D = D_1 \times \dots \times D_n$.

Then the causal decoder $d \in D$ is a sequence of functions $d = (d_i : 1 \leq i \leq n)$:

$$z_i = d_i(y^i) \quad (18)$$

Belief update:

In the above discussion on the causal decoder, we deliberately consider Z to be general, not necessarily equal to W . Indeed, as we shall see, in some cases we set $Z = \mathcal{P}(W)$ so that the outputs of the causal decoder represent *beliefs* about the source at time i . Define the beliefs $B_{i|j} \in \mathcal{P}(W)$ about the source at time i given the decoder's observations up until time $j \leq i$ as, for any $A \in \mathcal{B}(W)$:

$$B_{i|j}(A) \triangleq \mathbb{P}(W_i \in A | Y^j), \quad (19a)$$

$$b_{i|j}(A) \triangleq P_{W_i|Y^j=y^j}(A), \quad (19b)$$

The beliefs can be interpreted as state variables that can be updated sequentially given new observations. The *nonlinear filter* $\Lambda : \mathcal{P}(W) \times Y \times \tilde{E} \rightarrow \mathcal{P}(W)$ and *one step prediction update* $\Phi : \mathcal{P}(W) \rightarrow \mathcal{P}(W)$ rules are given by [52]:

$$\Lambda(b, y, \tilde{e})(dw) = \frac{dP_{Y|X}(\cdot | \tilde{e}(w))}{dP_\Lambda(\cdot | b, \tilde{e})}(y) \times \Phi(b)(dw) \quad (20)$$

$$\Phi(b)(dw) \triangleq \int_{w' \in W} Q_W(dw|w') b(dw') \quad (21)$$

$$P_\Lambda(dy|b, \tilde{e}) \triangleq \int_{w' \in W} P_{Y|X}(dy|\tilde{e}(w')) \Phi(b)(dw') \quad (22)$$

(20) can be interpreted as a standard manifestation of Bayes' rule: the numerator is simply the *likelihood*, the denominator is a *normalization constant*, and the coefficient $\Phi(b)$ is simply the *prior*. The aforementioned two equations specify how the beliefs are sequentially updated:

Lemma III.1 ([37],[52]). *For any i and encoder policy e_i with associated \tilde{e}_i given by (16), the following holds:*

$$b_{i|i-1} = \Phi(b_{i-1|i-1}) \quad (23a)$$

$$b_{i|i} = \Lambda(b_{i-1|i-1}, y_i, \tilde{e}_i) \quad (23b)$$

In Section IV, we demonstrate using a structural result how the beliefs arise as sufficient statistics in our main problem. In Section V, we demonstrate how they additionally serve as optimal decision variables with information gain cost (2).

Additive cost function:

Denote a coordination strategy, also termed policy, as $\pi = (e_1, \dots, e_n, d_1, \dots, d_n)$ and the set of all feasible policies as $\Pi = \{\pi : e_i \in E_i, d_i \in D_i\}$. The causal encoder and decoder e and d are cooperating to achieve a common goal. The performance of their cooperation is measured in terms of an expected sum of costs over time horizon n with the following structure:

$$J_{n,\pi}^\alpha = \mathbb{E}_\pi \left[\sum_{i=1}^n \rho(W_i, Z_{i-1}, Z_i) + \alpha \eta(X_i) \right] \quad (24)$$

The above expectation is taken with respect to an initial distribution P_{W_0, Z_0} where Z_0 is assumed known to the encoder and decoder. We assume that the functions ρ and η along with constant α have the following structure:

- $\rho : W \times Z \times Z \rightarrow \mathbb{R}_+$ is a ‘distortion-like’ source cost, that relates the distortion between the source at time i and the outputs in the vicinity of time i .
- $\eta : X \rightarrow \mathbb{R}_+$ is a ‘power-like’ channel input cost that penalizes channel inputs that deviate significantly from nominal desired values
- $\alpha \in \mathbb{R}^+$ balances the relative importance of the two costs.

Definition III.2. We say that a sequential encoder-decoder pair $\pi^* \in \Pi$ is (globally) optimal if

$$J_{n,\pi^*}^\alpha \leq J_{n,\pi}^\alpha \text{ for all } \pi \in \Pi. \quad (25)$$

IV. Main Structural Results

In this section, we prove that - under mild technical assumptions - for a general class of cost functions (ρ, η, α) inducing an average cost specified in (24), an optimal belief-based policy-estimator pair exists with the structure as shown in Fig 4.

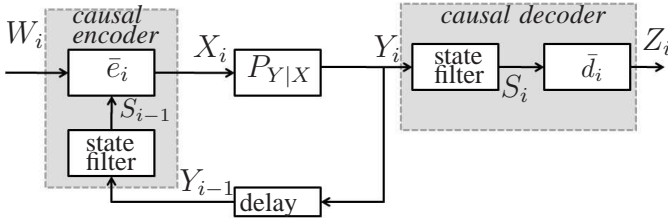


Fig. 4. Structural Result and Sufficient Statistics

We first consider the basic solution approach to the problem by first demonstrating an example with two time steps. The essence of the idea is as follows:

$$\begin{aligned}
 & \min_{e_1, d_1, e_2, d_2} \sum_{i=1}^2 \mathbb{E} [\rho(W_i, Z_{i-1}, Z_i) + \alpha \eta(X_i)] \\
 &= \mathbb{E} \left[\underbrace{\min_{e_1} \mathbb{E} [\alpha \eta(X_1) | \tilde{E}_1]}_{\text{controller at stage 0}} \right] \\
 &+ \mathbb{E} \left[\underbrace{\min_{e_2, d_1} \mathbb{E} [\rho(W_1, Z_0, Z_1) + \alpha \eta(X_2) | Z_0, Y^1, \tilde{E}_2, Z_1]}_{\text{controller at stage 1}} \right] \\
 &+ \mathbb{E} \left[\underbrace{\min_{d_2} \mathbb{E} [\rho(W_2, Z_1, Z_2) | Z_1, Y^2, Z_2]}_{\text{controller at stage 2}} \right] \quad (26a)
 \end{aligned}$$

Note that by grouping (d_i, e_{i+1}) in this manner, in stage i , $d_i : Y^i \rightarrow Z$ and $e_{i+1} : W^{i+1} \times Y^i \rightarrow X$ have access to a common piece of information, y^i (and thus also z_i). Note that $\tilde{e}_{i+1}(w^i, y^i)(\cdot) \equiv \tilde{e}_{i+1}(\cdot) : W \rightarrow X$ is a mapping as given by (16), whose alphabet, \tilde{E} , does not grow with i . We secondly consider the belief $b_{i|i}$, which is a function of (y^i, e_1, \dots, e_i) and whose alphabet, $\mathcal{P}(W)$, does not grow with i . The ‘control’ action taken by the encoder (decoder) are given by $\tilde{E}_{i+1}(Z_i)$ respectively.

We next demonstrate that the conditional expectations in (26) can be described in terms of these and other variables whose alphabets do not grow with i :

Lemma IV.1. For a fixed policy $\pi = (e, d)$, define $b' = b_{i|i} \in \mathcal{P}(W)$ as in (19), and $\tilde{e}_{i+1}(w^i, y^i)(\cdot) \triangleq \tilde{e}_{i+1}(\cdot) \in \tilde{E}$ as in (16). Define the state space $S = Z \times \mathcal{P}(W)$ and control space $U = \tilde{E} \times Z$ with $s_i \in S, u_i \in U$ given by

$$s_i = (z_{i-1}, b_{i|i}), \quad u = (\tilde{e}_{i+1}, z_i). \quad (27)$$

Then

$$\mathbb{E} [\rho(W_i, Z_{i-1}, Z_i) | Z_{i-1}, Y^i, Z_i] = \bar{\rho}(S_i, Z_i) \quad (28a)$$

$$\mathbb{E} [\eta(X_{i+1}) | Z_{i-1}, Y^i, \tilde{E}_{i+1}] = \bar{\eta}(S_i, \tilde{E}_{i+1}) \quad (28b)$$

To emphasize, this demonstrates state (s) and control (u) variables whose alphabets do not grow with i , for which ‘distortion’ and ‘cost’ like functions solely operate on. The definitions of $\bar{\rho}$ and $\bar{\eta}$ along with the lemma’s proof can be found in Appendix A. We now demonstrate that these state and control variables comprise a controlled Markov chain:

Lemma IV.2. The state $s_i = (z_{i-1}, b_{i|i})$ and control $u = (\tilde{e}_{i+1}, z_i)$ variables comprise a controlled Markov chain:

$$(a) \quad J_{n,\pi}^\alpha = \mathbb{E}_\pi \left[\sum_{i=0}^n \bar{g}_i(S_i, U_i) \right] \quad (29)$$

$$\bar{g}_i(s, u) \triangleq \begin{cases} \alpha \bar{\eta}(s_i, \tilde{e}_{i+1}) & i = 0 \\ \bar{\rho}(s_i, z_i) + \alpha \bar{\eta}(s_i, \tilde{e}_{i+1}) & 1 \leq i \leq n-1 \\ \bar{\rho}(s_i, z_i) & i = n \end{cases} \quad (30)$$

$$(b) \quad P_{S_{i+1}|S^i, U^i}(ds_{i+1}|s^i, u^i) = P_{S_{i+1}|S_i, U_i}(ds_{i+1}|s_i, u_i) \\ \triangleq Q_S(ds_{i+1}|s_i, u_i)$$

The proof of (a) follows directly from the law of iterated expectation and the definition (28). The proof of (b) can be found in Appendix B. Now define the cost-to-go function at stage $n-k$ as $V_{n-k} : S \rightarrow \mathbb{R}$. Then for $V_{n+1}(s) \equiv 0$ and $k = 1, \dots, n$ define:

$$V_{n-k}(s) = \inf_{u \in U} \left[\bar{g}_{n-k}(s, u) + \int_{s'} V_{n-k+1}(s') Q_S(ds'|s, u) \right] \quad (31)$$

This allows for us to state our main theorem of this section:

Theorem IV.3. If for each $s \in S$, the infimum in (31) is attained and the functions $(V_k : k = 0, \dots, n)$ are universally measurable, then there exists an optimal encoder/decoder policy (e^*, d^*) pair of the form

$$e_{i+1}^*(w^{i+1}, y^i) \equiv \tilde{e}_{i+1}^*(w_{i+1}, z_{i-1}, b_{i|i}) \quad (32a)$$

$$d_i^*(y^i) \equiv \bar{d}_i^*(z_{i-1}, b_{i|i}) \quad (32b)$$

Proof: Using standard dynamic programming arguments, [3, Chapter 8] we have that $J_{n,\pi^*}^\alpha \geq \mathbb{E}[V_0(S_0)]$. Next, $J_{n,\pi^*}^\alpha = \mathbb{E}[V_0(S_0)]$ and it can be implemented by a policy of the form (32) by a policy that attains the infimum of (31) for each s [3, Prop 8.6]. ■

The structural result in graphical form is shown in Figure 4. Note that within the causal encoder, the first process is a filter

that computes sufficient statistics. From here, these sufficient statistics are given to another encoder, \bar{e}_i , that uses them, along with the current source value, W_i , to specify the next channel input X_i . Analogously, the causal decoder is comprised of first the same recursive filter that computes sufficient statistics, followed by another decoder, \bar{d}_i , that computes Z_i .

We now note that ‘universal measurability’ [3] is usually satisfied:

Remark 1. *Standard technical assumptions guarantee universal measurability and that the infimum is attained; one example is as follows: (a) W , X , Y , and Z are compact Borel metric spaces, (b) ρ and η are lower semi-continuous, (c) $P_{Y|X}(dy|x)$ and $Q_W(dw|w')$ are continuous stochastic kernels, and (d) \tilde{E} is an equicontinuous space of functions.*

We also note our result generalizes the classical result of Walrand and Varaiya [37]:

Remark 2. *This result instantiates the result in [37] which assumes all alphabets are finite, $\eta \equiv 0$, and $\rho(w_i, z_{i-1}, z_i) \equiv \rho(w_i, z_i)$: (i) because of the finite alphabets and costs, the infimum is attained in (31); (ii) (27) can be collapsed to $s_i = b_{i|i}$ because of the absence of z_{i-1} in the function ρ . Secondly, our proof technique differs from [37, Sec. IV] in that we replace the three-step proof technique of ([37, Thm 1, Lemma 1, Thm 2]) - which includes two DP arguments ([37, Thm 1, Thm 2]) - with a single DP argument.*

However, our emphasis is not solely on allowing general alphabets or using the cost function of a particular form - both of these have in essence been accomplished using state augmentation and dynamic programming over general spaces. Rather, our emphasis is to carefully augment standard formulations to uncover an interplay information theory and control theory problems, as we shall see in the next section.

V. The Sequential Information Gain Cost

In this section, we specifically consider a class of problems that are not covered in traditional causal coding/decoding frameworks [37], [38],[39],[20, Ch. 6].

Traditional problems consider cost functions of the form $\rho(w_i, z_i)$ and assume that either all alphabets are finite [37], [38], or $W = Z = \mathbb{R}$ [39],[20, Ch. 6]. Motivated by the feedback communication example in Section I A, we now assume that $Z = \mathcal{P}(W)$, the space of possible beliefs on the source. Secondly, we construct $\rho(w_i, z_{i-1}, z_i)$ to be a log-likelihood ratio that is suggestive of an ‘information gain’-like quantity.

The following Lemma describes the relationship between $I(W^n; Y^n)$ and $I(W^n \rightarrow Y^n)$ for our problem setup (14)-(18). Because there is no feedback loop from Y to the generative process of W , these two quantities are equivalent:

Lemma V.1. *For any ‘sufficient statistic operating’ encoder $\pi \in \Pi$ satisfying (32a), i.e. $x_{i+1} = \bar{e}_{i+1}(w_{i+1}, z_{i-1}, b_{i|i})$, the following holds:*

$$I(W^n; Y^n) = I(W^n \rightarrow Y^n) = \sum_{i=1}^n I(W_i; Y_i | Y^{i-1}).$$

The proof is in Appendix C. Note that from the structural result in Theorem IV.3, there is no loss in performance for restricting attention to encoders of the form (32a). Under such encoders, note that the mutual information can be expressed as an accumulation of *sequential information gains*,

$$I(W^n; Y^n) = \sum_{i=1}^n I(W_i; Y_i | Y^{i-1}) \quad (33a)$$

$$= \sum_{i=1}^n \mathbb{E} [D(B_{i|i} \| B_{i|i-1})] \quad (33b)$$

$$= \sum_{i=1}^n \mathbb{E} \left[\log \frac{dB_{i|i}}{d\Phi(B_{i-1|i-1})}(W_i) \right] \quad (33c)$$

where (33a) follows from Lemma V.1; (33b) follows from (9) and (19); and (33c) follows from (6) and (21).

One may consider finding encoder policies e in order to maximize $I(W^n; Y^n)$, using a state space approach over the space of beliefs. [19] formulated a stochastic control problem where $B_{i-1|i-1}$ is a state variable and the only decision variable is the causal encoder’s strategy - the decoder did not specify a decision variable Z_i . There, it was shown that when W is uniformly distributed on $W = [0, 1]$ and $(W_i = W : i \geq 1)$, the causal encoder given by the posterior matching scheme by Shayevitz and Feder [15] is an optimal solution to a control problem where costs are related to conditional mutual informations (33b). Anand and Kumar have recently considered a related problem where [36] where $(W_i : i \geq 1)$ is a general Markov process over a *finite* alphabet, and the cost function is a conditional mutual information. There, also, however, the decoder did not specify a decision variable Z_i .

In this setting, we do not treat $B_{i|i}$ as a state variable; rather, we first consider a problem in the framework of causal coding/decoding, where the decoder’s decision variable Z_i can be *any* possible belief: $Z = \mathcal{P}(W)$. In order to reward larger information gains, we define an appropriate cost pertaining to the negative logarithm of the Radon-Nikodym derivative evaluated at w_i that is inspired by the expansion of mutual information given in (33c):

$$\rho(w_i, z_{i-1}, z_i) = \begin{cases} -\log \frac{dz_i}{d\Phi(z_{i-1})}(w_i) & \text{if } z_i \ll \Phi(z_{i-1}) \\ \infty, & \text{otherwise} \end{cases} \quad (34)$$

The reason we assign $\rho = \infty$ when $z_i \ll \Phi(z_{i-1})$ is because under any reasonable belief-setting strategy, if the belief about W_i given Y^{i-1} - the one-step prediction update (21) given by $\Phi(z_{i-1})$ - assigns zero probability mass to $A \in \mathcal{B}(W)$, then the belief about W_i given Y^i - which is given by Z_i should also.

We emphasize here that the beliefs on the source are themselves decision variables, which are what the causal decoder must specify. This viewpoint has been used within the sequential prediction literature [28] and statistical signal processing [30] but appears to not have been used as frequently in the literature that attempts to draw synergies between information theory and control.

Define $Z_0(A) = \mathbb{P}(W_0 \in A)$, the distribution on W_0 . We now state the following useful Lemma that decomposes the

cost into the state and distortion parts, that act on different aspects of the control input:

Lemma V.2. *Under the information gain criterion (34), for a state variable $s_i = (z, b)$ and control variable $u_i = (\tilde{e}, z')$,*

$$\bar{\eta}(s_i, \tilde{e}) = \int_{w \in W} \eta(\tilde{e}(w)) \Phi(b)(dw) \quad (35)$$

$$\bar{\rho}(s_i, z') = \begin{cases} D(b\|z') - D(b\|\Phi(z)) & b \ll z' \ll \Phi(z) \\ \infty & \text{otherwise} \end{cases} \quad (36)$$

The proof can be found in Appendix D.

With this, we state the main theorem of our section. It says that when treating beliefs as decision variables, under the information gain criterion (34), the optimal decision rule for the decoder is to select its belief about W_i to be $z_i = b_{i|i}$, and the optimal decision rule for the encoder is to maximize mutual information subject to a cost on channel inputs:

Theorem V.3. *Under cost criterion (36), there exists an optimal encoder/decoder policy (e^*, d^*) pair of the form*

$$e_{i+1}^*(w^{i+1}, y^i) \equiv \bar{e}_{i+1}^*(w_{i+1}, b_{i|i}) \quad (37)$$

$$d_i^*(y^i) \equiv \bar{d}_i^*(b_{i|i}) = b_{i|i} \quad (38)$$

where $b_{i|i} = \Lambda(b_{i-1|i-1}, y_i, \bar{e}_i^*(\cdot, b_{i-1|i-1}))$ and the optimal cost is given by

$$J_{n, \pi^*}^\alpha = \min_{e \in \mathcal{E}} -I(W^n; Y^n) + \alpha \mathbb{E}_e \left[\sum_{i=1}^n \eta(X_i) \right]. \quad (39)$$

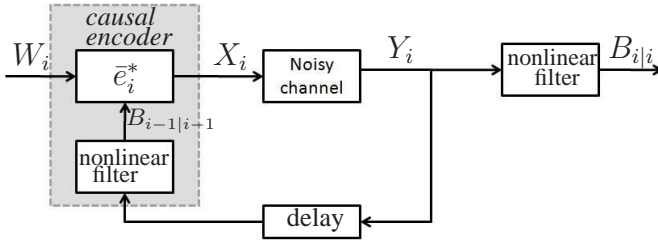


Fig. 5. Simplified structural result with $Z = \mathcal{P}(W)$ and sequential information gain cost (36).

The proof can be found in Appendix E.

Remark 3. *The proof of Theorem V.3 (Appendix E) uses dynamic programming, the second law of thermodynamics for Markov chains, and exploits how the divergence acts as a Lyapunov function for the stability of the nonlinear filter. This further demonstrates an interesting relationship between information theory and thermodynamics [31], [32]. This idea of using beliefs as decision variables where the posterior belief is optimal has been used in sequential prediction [28] and in variational approaches to nonlinear estimation [31], but within the context of causal coding decoding problems, this is to the best of our knowledge, new.*

We will demonstrate in the examples section how this relates to the hidden Markov model and the nonlinear filter as well as the Posterior Matching Scheme [15] for communication of a message point over a noisy channel with feedback.

VI. Inverse Optimal Control with Stationary Markov Coordination Strategies

In the last section, we demonstrated that for a specific “information-gain” related cost function (34), there existed an optimal encoder policy of the form $X_i = \bar{e}_i^*(W_i, Z_{i-1})$ and decoder policy of the form $Z_i = B_{i|i} = \Lambda(B_{i-1|i-1}, Y_i, \bar{e}_i^*(\cdot, B_{i-1|i-1})) = \bar{d}(Z_{i-1}, Y_i)$.

In light of this, we now consider a general Markov process $W \in \mathcal{W}$ and $Z \in \mathcal{Z}$ where Z need not be $\mathcal{P}(W)$ and fix the coordination strategy $\bar{\pi}$ to be *stationary Markov (SM)*, meaning that for fixed functions $\bar{e} : \mathcal{W} \times \mathcal{Z} \rightarrow \mathcal{X}$ and $\bar{d} : \mathcal{Z} \times \mathcal{Y} \rightarrow \mathcal{Z}$, the following holds:

$$x_i = \bar{e}(w_i, z_{i-1}) \quad (40a)$$

$$z_i = \bar{d}(z_{i-1}, y_i). \quad (40b)$$

See Figure 6. (40b) is sometimes termed a decoder of ‘finite-memory’ [37],[60]. Since the encoder and decoder both utilize z_{i-1} , this can be also interpreted as a collection of ‘equi-memory’ encoders and decoders [20, Definition 6.3.2]. For

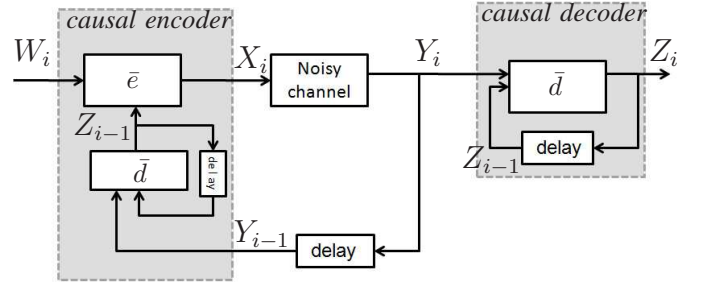


Fig. 6. A Stationary Markov coordination strategy

a fixed SM coordination strategy $\bar{\pi}$, we compare it against arbitrary policies of the form $\pi = (e_1, \dots, e_n, d_1, \dots, d_n)$ where $e_i : \mathcal{W}^i \times \mathcal{Y}^{i-1} \rightarrow \mathcal{X}$ is given in (15), and $d_i : \mathcal{Y}^i \rightarrow \mathcal{Z}$ is given in (18). Here we identify the structure of cost functions $\rho(w_i, z_{i-1}, z_i)$ under which $\bar{\pi}$ is globally optimal.

Definition VI.1. *A coordination strategy π is inverse-control optimal for a source-channel pair $(P_{W^n}, P_{Y|X})$ if $J_{n, \pi}^\alpha \leq J_{n, \pi'}^\alpha$ for all $\pi' \in \Pi$ for some $\alpha \geq 0$, $\rho : \mathcal{W} \times \mathcal{Z} \times \mathcal{Z} \rightarrow \mathbb{R}$ and $\eta : \mathcal{X} \rightarrow \mathbb{R}_+$ in (24).*

To develop high-level conditions under which π is indeed inverse-control optimal, we first develop some preliminary machinery. Fix a specific ρ and η function. For any coordination strategy π , define $P_{Z^n|W^n}^\pi$ as the conditional distribution induced statistical law under π and also define:

$$D_\pi \triangleq \frac{1}{n} \mathbb{E}_\pi \left[\sum_{i=1}^n \rho(W_i, Z_{i-1}, Z_i) \right] \quad (41)$$

$$L_\pi \triangleq \frac{1}{n} \mathbb{E}_\pi \left[\sum_{i=1}^n \eta(X_i) \right] \quad (42)$$

Define the rate-distortion function for P_{W^n} and ρ as [2]

$$R_n(\rho, P_{W^n}, D) \triangleq \min_{P_{Z^n|W^n} : \mathbb{E} \left[\frac{1}{n} \sum_{i=1}^n \rho(W_i, Z_{i-1}, Z_i) \right] \leq D} \frac{1}{n} I(P_{W^n}, P_{Z^n|W^n}) \quad (43)$$

and denote $P_{Z^n|W^n}^*(\rho, P_{W^n}, D)$ as the minimizer in (43). We now state the following standard lemma:

Lemma VI.2. Fix a $\pi, P_{W^n}, P_{Y|X}, \rho$, and η . Then

$$R_n(\rho, P_{W^n}, D_\pi) \leq \frac{1}{n} I(P_{W^n}, P_{Z^n|W^n}^\pi) \quad (44a)$$

$$\leq \frac{1}{n} I(P_{W^n}, P_{Y^n|W^n}^\pi) \quad (44b)$$

$$\leq \frac{1}{n} \sum_{i=1}^n I(P_{X_i}^\pi, P_{Y|X}) \quad (44c)$$

$$\leq C(\eta, P_{Y|X}, L_\pi) \quad (44d)$$

where equality holds if and only if:

- (a) $P_{Z^n|W^n}^\pi = P_{Z^n|W^n}^*(\rho, P_{W^n}, D_\pi)$
- (b) $I(P_{W^n}, P_{Z^n|W^n}^\pi) = I(P_{W^n}, P_{Y^n|W^n}^\pi)$
- (c) $I(Y_i; Y^{i-1}) = 0$ for each i
- (d) $P_{X_i}^\pi = P_X^*(\eta, P_{Y|X}, L_\pi)$ for each i

The proof is standard [2] but for the sake of completeness, we include it in Appendix F. This leads to an intermediate sufficient condition for inverse control optimality that applies for any $\pi \in \Pi$ (e.g. π need not be stationary-Markov):

Lemma VI.3. If a policy $\pi \in \Pi$ results in (44) holding with equality, then it is inverse control optimal.

Proof: Note that $J_{n,\pi}^\alpha = D_\pi + \alpha L_\pi = \langle (1, \alpha), (D_\pi, L_\pi) \rangle$. Define $\mathcal{R} = \{(D_{\pi'}, L_{\pi'} : \pi' \in \Pi)\}$. Define $\tilde{\Pi}$ to be the set of randomized policies in Π . Note that any $\pi' \in \tilde{\Pi}$ still induces a conditional distribution $P_{Z^n|W^n}^{\pi'}$ and thus an induced $D_{\pi'}$ and $L_{\pi'}$ so that we may define $\tilde{\mathcal{R}} = \{(D_{\pi'}, L_{\pi'} : \pi' \in \tilde{\Pi})\}$. Clearly, $\mathcal{R} \subset \tilde{\mathcal{R}}$, and secondly, $\tilde{\mathcal{R}}$ is convex. Next, note that if (44) holds with equality for some $\pi \in \Pi$, then $D_\pi \leq D_{\pi'}$ for any $\pi' \in \tilde{\Pi}$ for which $L_{\pi'} \leq L_\pi$ from the definition of $R_n(\rho, P_{W^n}, D_\pi)$ in (43) and $C(\eta, P_{Y|X}, L_\pi)$ in (11) (See also [4, Lemma 1]). Thus (D_π, L_π) is a boundary point of $\tilde{\mathcal{R}}$. Therefore there exists a supporting hyperplane parametrized by $\alpha \geq 0$ that intersects (D_π, L_π) :

$$J_{n,\pi}^\alpha = \langle (1, \alpha), (D_\pi, L_\pi) \rangle \leq \langle (1, \alpha), (D_{\pi'}, L_{\pi'}) \rangle = J_{n,\pi'}^\alpha.$$

for all $\pi' \in \tilde{\Pi} \supset \Pi$ [61], where $\langle \cdot, \cdot \rangle$ denotes inner product. ■

We now consider SM policies for which condition (c) in Lemma VI.2 holds, and demonstrate a stationary Markov relationships between (W^n, Z^n) random variables:

Lemma VI.4. If an SM coordination strategy $\bar{\pi}$ (40) induces the channel outputs $(Y_i : 1 \leq i \leq n)$ being i.i.d., then

$$P_{Z_i|Z^{i-1}=z^{i-1}, W^n=w^n}(dz_i) \equiv Q_{Z'|Z, W'}^{\bar{\pi}}(dz_i|z_{i-1}, w_i) \quad (45a)$$

$$P_{Z_i|Z^{i-1}=z^{i-1}}(dz_i) \equiv Q_{Z'|Z}^{\bar{\pi}}(dz_i|z_{i-1}) \quad (45b)$$

The proof of this can be found in Appendix G and exploits the equivalence between a random process being a time-homogeneous Markov chain and it being represented as an iterated function system [62]. With this, we can now state the main theorem of this section:

Theorem VI.5. If under a SM policy $\bar{\pi}$, $(Y_i : 1 \leq i \leq n)$ are i.i.d. and $I(P_{W^n}, P_{Z^n|W^n}^{\bar{\pi}}) = I(P_{W^n}, P_{Y^n|W^n}^{\bar{\pi}})$, then $\bar{\pi}$ is inverse control optimal with ρ and η given by:

$$\eta(x) \propto_+ D(P_{Y|X=x} \| P_Y^{\bar{\pi}}) \quad (46a)$$

$$\rho(w_i, z_{i-1}, z_i) \propto_+ -\log \frac{dQ_{Z'|Z, W'}^{\bar{\pi}}(\cdot|z_{i-1}, w_i)}{dQ_{Z'|Z}^{\bar{\pi}}(\cdot|z_{i-1})}(z_i) \quad (46b)$$

where \propto_+ denotes proportional to with a positive constant.

Proof: Note that it suffices to show that (44) holds with equality and then invoke Lemma VI.3. First note that from the theorem definition, clearly conditions (b) in Lemma VI.2 holds with equality. Since $(Y_i : i = 1, \dots, n)$ are i.i.d. and since the channel is memoryless, it follows that the $(X_i : i = 1, \dots, n)$ are identically distributed and so condition (c) in Lemma VI.2 holds with equality. Thus the two remaining conditions are to show that conditions (a) and (d) in Lemma VI.2 hold with equality.

The variational equations for an optimal solution to (43) state that a necessary and sufficient condition for $P_{Z^n|W^n}^* = P_{Z^n|W^n}^*(\rho, P_{W^n}, D_\pi)$ is the following relationship [47]:

$$\frac{dP_{Z^n|W^n=w^n}}{dP_{Z^n}^{\bar{\pi}}}(z^n) = \zeta(w^n) e^{-\alpha_2(\sum_{i=1}^n \rho(w_i, z_{i-1}, z_i))} \quad (47)$$

For our case, note that

$$\begin{aligned} \log \frac{dP_{Z^n|W^n=w^n}}{dP_{Z^n}^{\bar{\pi}}}(z^n) &= \sum_{i=1}^n \log \frac{dP_{Z_i|Z^{i-1}=z^{i-1}, W^n=w^n}}{dP_{Z_i|Z^{i-1}=z^{i-1}}}^{\bar{\pi}}(z_i) \\ &= \sum_{i=1}^n \log \frac{dQ_{Z'|Z, W'}^{\bar{\pi}}(\cdot|z_{i-1}, w_i)}{dQ_{Z'|Z}^{\bar{\pi}}(\cdot|z_{i-1})}(z_i) \end{aligned} \quad (48)$$

where (48) follows from Lemma VI.4. Thus we see that with ρ given by (46b), from (47) we see that condition (a) of Lemma 44 holds with equality.

Lastly, condition (d) of Lemma 44 holds with equality if and only if each $P_{X_i}^{\bar{\pi}} \sim P_X^*(\eta, P_{Y|X}, L_\pi)$. Variational arguments [4, Lemma 3], [48, p. 147] demonstrate that this criterion is equivalent to (46a). ■

Corollary VI.6. if the function $\bar{d}(z_{i-1}, \cdot) \triangleq \bar{d}_{z_{i-1}}(\cdot)$ in (40b) is invertible, then condition (46b) in Theorem VI.5 becomes

$$\rho(w_i, z_{i-1}, z_i) \propto_+ \log \frac{dP_{Y|X=\bar{e}(w_i, z_{i-1})}}{dP_Y^{\bar{\pi}}}\left(\bar{d}_{z_{i-1}}^{-1}(z_i)\right)$$

We now first relate this to ‘source-channel’ matching and how it is in some sense it is also ‘natural’ within the inverse control framework to have a distortion function of the form $\rho(w_i, z_{i-1}, z_i)$:

Remark 4. The problem setup leading up to Theorem VI.5 is philosophically inspired by the ‘source channel matching’ work [4], [49] - but here, we are relating this to a causal coding-decoding problem with **causal encoder feedback**, and **time-invariant additive costs**. These two properties appear to make the distortion function $\rho(w_i, z_{i-1}, z_i)$ - as compared to $\rho(w_i, z_i)$ - crucially important: note the time-invariant statistical relationships in Lemma VI.4 and how they relate to $\rho(w_i, z_{i-1}, z_i)$ in (48) and Corollary VI.6 pertaining to condition (a) in Lemma VI.2. With this more general $\rho(w_i, z_{i-1}, z_i)$

framework, we can characterize time-invariant cost functions for problems where neither W nor Z are stationary (see the linear quadratic Gaussian decentralized control and M/M/1 queue examples in Section VII).

Next, we demonstrate how time-reversibility of an appropriately defined Markov chain can serve for Theorem VI.5 - and thus inverse control optimality - to hold for SM coordination strategies.

A. Time-Reversibility of Markov Chains and Inverse Optimal Control

Time reversibility plays an important role in disciplines concerning dynamical systems, e.g. in physics (conservation laws); statistical mechanics (in terms of equilibrium states); stochastic processes (e.g. queuing networks [63], [64] and convergence rates of Markov chains [65, ch 20]); and biology (e.g. trans paths in ion channels [66]). However, its use in acting as a sufficient condition to saturate fundamental information-theoretic limits appears to be somewhat limited. One special noteworthy exception is how Mitter and colleagues have related Markov chain reversibility to rate of entropy production in non-equilibrium thermodynamics [51], [32, Remark 2.1].

In queuing systems, the celebrated Burke's theorem [64], [67] uses Markov chain time reversibility to show that, in a certain stochastic dynamical system - an M/M/1 queue in steady-state - the state of the system (queue) at time i is statistically independent of all outputs (departures) before time i . This observation has been used in proving achievability theorems using for queuing timing channels [54], [55], [68], and for implementing recursive schemes that maximize mutual information according to the converse to the channel coding theorem with feedback [16], [17], [18], [15]. Here we demonstrate how time reversibility of Markov chains provides a sufficient condition for inverse optimal control with SM coordination strategies.

We first note that from (14), W is a time-homogenous Markov chain and so it can be represented as an iterated function system [62]:

$$W_i = \psi(\tilde{W}_i, W_{i-1}) \equiv \psi_{\tilde{W}_i}(W_{i-1}), \quad i \geq 1 \quad (49)$$

where \tilde{W}_i are i.i.d. To ensure, (13), we assume

$$I(\tilde{W}_i; X^{i-1}, Y^{i-1}) = 0. \quad (50)$$

We next suppose the structure of the SM coordination strategy is such that the following assumption holds

Definition VI.7. We say that the SM coordination strategy $\bar{\pi} = (\bar{e}, \bar{d})$ elicits 'reversibly feasible dynamics' if $I(P_{W^n}, P_{Z^n|W^n}^\pi) = I(P_{W^n}, P_{Y^n|W^n}^\pi)$ and the statistical dynamics can be described as

$$X_i = f(\tilde{X}_{i-1}, \tilde{W}_i) \equiv f_{\tilde{X}_{i-1}}(\tilde{W}_i) \quad (51)$$

$$\tilde{X}_i = g(X_i, Y_i) \equiv g_{X_i}(Y_i) \quad (52)$$

where $f_{\tilde{X}_{i-1}} : W \rightarrow X$ and $g_{X_i} : Y \rightarrow X$ are \mathbb{P} -a.s. invertible functions for $i = 1, \dots, n$.

Note that \tilde{x} in condition (52) is the update to the state *after* the output of the channel is taken into consideration and *before* the source w is updated to the state.

We now show an example that is related to feedback communication with posterior matching [15]

Example 1. Let $W = X = [0, 1]$ and $Z = \mathcal{P}(W)$. Then the 'posterior matching' scheme [15] given by

$$\tilde{W}_0 \sim \text{unif}[0, 1], \quad \tilde{W}_i \equiv 1, i \geq 1 \quad (53a)$$

$$W_0 = \tilde{W}_0, \quad W_i = W_{i-1}, i \geq 1 \quad (53b)$$

$$Z_i = B_{i|i} \quad (53c)$$

$$X_0 = 0, \quad X_i = \tilde{X}_{i-1}, i \geq 1 \quad (53d)$$

$$\tilde{X}_i = Z_{i-1}([0, W_i]) = F_{X|Y}(X_i|Y_i) \quad (53e)$$

elicits reversibly feasible dynamics. Note that this clearly is a SM coordination policy because for the decoder $B_{i|i}$ is given by the nonlinear filter, and for the encoder, this follows from the first equality in (53e). To verify that the last equality in (53e) holds, see [15, Corollary 6].

Next, we consider a SM coordination strategy that results in a birth-death Markov chain [63], [64] where X can increase or decrease by at most 1 from time i to time $i+1$ (see Figure 7):

Example 2. Let $\tilde{W} = W = X = Y = Z = \mathbb{F}$ for some field. Then the following SM coordination strategy

$$W_i = W_{i-1} + \tilde{W}_i \quad (54a)$$

$$Z_i = Z_{i-1} + Y_i \quad (54b)$$

$$\tilde{X}_i = X_i - Y_i \quad (54c)$$

$$X_i = W_i - Z_{i-1} = \tilde{X}_{i-1} + \tilde{W}_i \quad (54d)$$

elicits reversibly feasible dynamics. This follows from inspection.

See Figure 7.

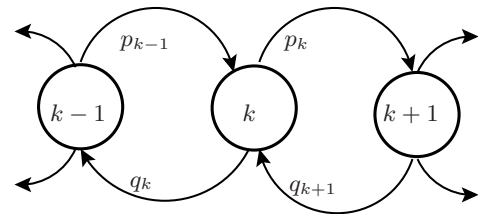


Fig. 7. A birth-death Markov chain X .

Lemma VI.8. Consider an SM coordination strategy with dynamics given by (54) where $\mathbb{P}(\tilde{W}_i \in \{0, 1\}) = \mathbb{P}(Y_i \in \{0, 1\})$. If X is a time-reversible Markov chain, then (X, \tilde{X}) is jointly a time-reversible Markov chain, Y_i are i.i.d., and $\bar{\pi}$ is inverse-control optimal.

The proof that (X, \tilde{X}) is jointly stationary and that Y_i are i.i.d. is a generalization [69] of the discrete-time Burke's theorem [67] from queuing theory. From there, the Lemma follows by simply invoking Definition VI.1 and Theorem VI.5. As of now, time-reversibility is only wed to inverse control optimality in the algebraic setup of Example 2. This alludes

to there being a more general statement under which time-reversibility implies inverse optimality:

Lemma VI.9. *If an SM coordination strategy $\bar{\pi}$ elicits reversibly feasible dynamics and (X, \tilde{X}) is jointly a time-reversible Markov chain, then $\bar{\pi}$ is inverse-control optimal.*

Proof: We now develop a generalization to the discrete-time proof of Burke's theorem [67], [69] from queuing theory. Note that from Assumption VI.7 that

$$\bar{W}_i = f_{\tilde{X}_{i-1}}^{-1}(X_i) \quad (55)$$

$$Y_i = g_{X_i}^{-1}(\tilde{X}_i) \quad (56)$$

Now note from the time-reversibility assumption, we have that

$$(X_1, \tilde{X}_1, \dots, \tilde{X}_{i-1}, X_i) \stackrel{d}{=} (X_{2i-1}, \tilde{X}_{2i-2}, \dots, \tilde{X}_i, X_i) \quad (57)$$

Re-arranging terms, we have

$$(X_i, \tilde{X}^{i-1}, X^{i-1}) \stackrel{d}{=} (X_i, \tilde{X}_i^{2i-2}, X_{i+1}^{2i-1}) \quad (58)$$

$$\Rightarrow I(X_i; \tilde{X}^{i-1}, X^{i-1}) = I(X_i; \tilde{X}_i^{2i-2}, X_{i+1}^{2i-1}) \quad (59)$$

$$\Leftrightarrow I(X_i; Y^{i-1}, X^{i-1}) = I(X_i; \tilde{X}_i^{2i-2}, \tilde{W}_{i+1}^{2i-1}) \quad (60)$$

where (60) follows from invariance of mutual information to bijective transformations and (51)-(52), (55)-(56); Analogously, from (58),

$$I(X_i; X^{i-1} | \tilde{X}^{i-1}) = I(X_i; X_{i+1}^{2i-1} | \tilde{X}_i^{2i-2}) \quad (61)$$

$$\Leftrightarrow I(X_i; X^{i-1} | Y^{i-1}) = I(X_i; \tilde{W}_{i+1}^{2i-1} | \tilde{X}_i^{2i-2}) \quad (62)$$

Therefore

$$I(X_i; Y^{i-1}) = I(X_i; \tilde{W}_{i+1}^{2i-1}) \quad (63)$$

$$= 0 \quad (64)$$

where (63) follows from the chain rule of mutual information (10) and subtracting (62) from (60); and (64) follows from (50). Because of the nature of the memoryless channel $P_{Y|X}$ in (17), it follows that $I(Y_i; Y^{i-1}) = 0$ for all i . Moreover, because Markov chain reversibility implies stationarity, it follows that $(Y_i : 1 \leq i \leq n)$ are i.i.d. Thus we can invoke Theorem VI.5. ■

B. Information Gain Cost and Inverse Optimal Control

In the beginning of this section, we motivated the definition of stationary Markov coordination strategies, of the type $x_i = \bar{e}_i(w_i, z_{i-1})$ and $z_i = \bar{d}_i(z_{i-1}, y_i)$ by noting from Section V that such an optimal decoder exists when $Z = \mathcal{P}(W)$ and the cost function is of the “information-gain” related structure (34):

$$X_i = \bar{e}_i^*(W_i, B_{i-1|i-1}) = \bar{e}_i(W_i, Z_{i-1})$$

$$Z_i = B_{i|i} = \Lambda(B_{i-1|i-1}, Y_i, \bar{e}_i^*(\cdot, B_{i-1|i-1})) = \bar{d}_i(Z_{i-1}, Y_i)$$

We now demonstrate that the information gain cost function in Section V can be seen to be a consequence of our inverse optimal control framework for any coordination strategy for $(Y_i : 1 \leq i \leq n)$ are i.i.d. and \bar{d} is the nonlinear filter:

Lemma VI.10. *Let $Z = \mathcal{P}(W)$. If a SM coordination strategy $\bar{\pi}$ contains a nonlinear filter decoder $z_i = \bar{d}(z_{i-1}, y_i) =$*

$\Lambda(z_{i-1}, y_i, \bar{e}(\cdot, y_{i-1}))$ and $(Y_i : 1 \leq n)$ are i.i.d., then $\bar{\pi}$ is inverse control optimal with information gain distortion $\rho(w_i, z_{i-1}, z_i) = -\log \frac{dz_i}{d\Phi(z_{i-1})}(w_i)$ and state cost function $\eta(x) = D(P_{Y|X=x} \| P_Y^{\bar{\pi}})$. The optimal cost is given by

$$J_{n,\bar{\pi}}^\alpha = (\alpha - 1)nC(\eta, P_{Y|X}, L_\pi). \quad (65)$$

Proof: First note that under this policy $\bar{\pi}$, $Z_i = B_{i|i}$. Now note that clearly

$$\mathbb{P}(W_i \in A | Y^i) = Z_i(A) = \mathbb{P}(W_i \in A | Z_{i-1}, Z_i), \quad (66a)$$

$$\mathbb{P}(W_i \in A | Y^{i-1}) = \Phi(Z_{i-1})(A) = \mathbb{P}(W_i \in A | Z_{i-1}) \quad (66b)$$

As such, we have that,

$$\begin{aligned} \frac{dz_i}{d\Phi(z_{i-1})}(w_i) &= \frac{dP_{W_i|Y^i}^{\bar{\pi}}(w_i)}{dP_{W_i|Y^{i-1}}^{\bar{\pi}}(w_i)} \\ &= \frac{dP_{W_i|Z_{i-1}=z_{i-1}, Z_i=z_i}^{\bar{\pi}}(w_i)}{dP_{W_i|Z_{i-1}=z_{i-1}}^{\bar{\pi}}(w_i)} \quad (67) \end{aligned}$$

$$= \frac{dP_{Z_i|Z_{i-1}=z_{i-1}, W_i=w_i}^{\bar{\pi}}(z_i)}{dP_{Z_i|Z_{i-1}=z_{i-1}}^{\bar{\pi}}(z_i)} \quad (68)$$

$$= \frac{dQ_{Z'|Z,W}^{\bar{\pi}}(\cdot | z_{i-1}, w_i)}{dQ_{Z'|Z}^{\bar{\pi}}(\cdot | z_{i-1})}(z_i)(z_i) \quad (69)$$

where (67) follows from (66); (68) follows from a simple application of Bayes' rule: $\frac{\mathbb{P}(A|B,C)}{\mathbb{P}(A|B)} = \frac{\mathbb{P}(C|A,B)}{\mathbb{P}(C|B)}$; and (69) follows from Lemma VI.4. Also, since $Z_i = B_{i|i}$, it follows that $I(W^n; Y^n) = I(W^n; Z^n)$. Thus Theorem VI.5 applies and so $\bar{\pi}$ is inverse control optimal. To characterize the final cost, note that for the associated α ,

$$J_{n,\bar{\pi}}^\alpha = -I(W^n; Y^n) + \alpha \mathbb{E}_{\bar{e}} \left[\sum_{i=1}^n \eta(X_i) \right] \quad (70)$$

$$= -nC(\eta, P_{Y|X}, L) + \alpha \mathbb{E}_{\bar{e}} \left[\sum_{i=1}^n \eta(X_i) \right] \quad (71)$$

$$= -nC(\eta, P_{Y|X}, L) + \alpha \left(\sum_{i=1}^n I(X_i; Y_i) \right) \quad (72)$$

$$= (\alpha - 1)nC(\eta, P_{Y|X}, L) \quad (73)$$

where (70) follows from Theorem V.3; (71) follows from the fact that Theorem VI.5 applies which means that (44) holds with equality; (72) follows from the definition of mutual information and that $\eta(x) = D(P_{Y|X=x} \| P_Y^{\bar{\pi}})$; and (73) follows from the fact that (44) holds with equality. ■

Traditionally, inverse optimal control is performed through finding a control-Lyapunov function (40), which involves performing a sequential decomposition of the problem and finding a consistent value function (31). When $Z = \mathcal{P}(W)$, this can be done using only the decision variables and stationary-Markov coordination strategies as in Section V: $X_i = \bar{e}_i(W_i, Z_{i-1})$ and $Z_i = B_{i|i} = \Lambda(B_{i-1|i-1}, Y_i, \bar{e}_i(\cdot, B_{i-1|i-1})) = \bar{d}(Z_{i-1}, Y_i)$. This means that using a control-Lyapunov approach, first a sequential decomposition resting upon the structural result work in Section IV would be needed, with the additional effort of showing that coordination strategies of the structural result form (32) can be reduced to stationary Markov strategies of the form (40). However, our inverse

control optimality sufficient conditions apply for a general Z (which need not be $\mathcal{P}(Z)$) and do not involve a sequential decomposition. As such, the approach developed in this section - when applicable - appears to require ‘less effort’ than typically required in arriving at an inverse optimal control result.

VII. Examples

In this section, we provide examples of the Theorems and Lemmas from previous sections.

A. Likelihood Ratio Cost and Information Gain: HMMs and the Nonlinear Filter

We now demonstrate that the information gain cost framework of Section V demonstrates the causal coding/decoding and information-theoretic optimality of the nonlinear filter in a specific sense. Related work on using variational principles to characterize the nonlinear filter was reported in [31]. However, demonstrating that the nonlinear filter is acting as an optimal controller with respect to this information gain cost function, is - to the best of our knowledge - new. We start by considering the following assumptions:

- (i) the source and channel inputs have the same alphabets: $W = X$
- (ii) the causal encoder alphabet $E_i = \{e_i : W^i \times Y^{i-1} \rightarrow X\} = \{=\}$ where $=$ is the identity function: $x_i = w_i$.

Under these conditions, the only feasible encoder simply specifies w_i as the channel inputs, and thus this becomes a hidden Markov model.

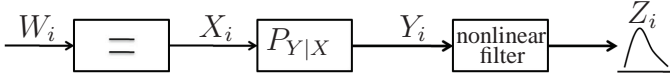


Fig. 8. The information gain cost when the encoder set consists of only the identity function. This becomes a hidden Markov model where the nonlinear filter is an optimal solution.

As such, we can consider maximizing the mutual information from W^n to Y^n over all possible causal decoder policies. As such, the optimal design of e_i disappears and the focus becomes optimal design of $\{d_i\}$. We now show that, assuming $Z_0(A) = \mathbb{P}(W_0 \in A)$, the optimal policy for the decoder is given by the true posterior - which can be computed recursively using the nonlinear filter:

Lemma VII.1. *Under assumptions (i) and (ii) above, and cost functions $\eta \equiv 0$ and ρ given by (34)*

$$\rho(w_i, z_{i-1}, z_i) = \begin{cases} -\log \frac{dz_i}{d\Phi(z_{i-1})}(w_i) & \text{if } z_i \ll \Phi(z_{i-1}) \\ \infty, & \text{otherwise} \end{cases},$$

the policy $\bar{\pi}$ consisting of the identity function encoder and nonlinear filter decoder $Z_i = B_{i|i} = \Lambda(Z_{i-1}, Y_i, =)$, is globally optimal where $J_{n,\pi^*}^\alpha = -I(W^n; Y^n)$.

Proof: Because E_i is a singleton consisting of the identity function, and because $\eta(x) = 0$, this follows directly from Theorem V.3. ■

B. Likelihood Ratio Cost and Information Gain: Feedback Communication of a Message Point

Given that the natural mathematical framework to handle feedback is control theory, we consider the problem of communication over noisy channels with feedback from the dynamical systems perspective, and make use of recent sequential approaches to communication. This viewpoint has been made largely possible by a recent development in the information theory literature - the posterior matching (PM) scheme [15] - which generalizes other ‘message-point’ style feedback communication schemes [17], [18], [16]: rather than nR bits, a message point on the interval $[0, 1]$ is considered. The notion of “decoding nR bits” now becomes equivalent to determining the message point within an interval of length 2^{-nR} at the receiver (see Section I-A).

The implementational details and fundamental limits are completely in line with traditional communication paradigms (see [15]) but there are subtle, yet striking differences. Because the message point is a point on the $[0, 1]$ line, there is no pre-specified block length; the system operates to sequentially give the user the information that is “still missing” at the receiver. Moreover, at each time step, the decoder specifies an output $Z_i \in \mathcal{P}(W)$, which is a belief about the message point. We now demonstrate how this notion of communication, and the problem of finding the optimal encoder with feedback, can be captured with our framework. Moreover, we will demonstrate that the PM scheme is an optimal solution to the problem.

Let $W = [0, 1]$ and $Z = \mathcal{P}(W)$. Further, let the source process be the ‘repetition’ Markov process ($W_i = W : i \geq 1$) with W uniformly distributed over $[0, 1]$. If we assume that there is an expected cost constraint $\frac{1}{n} \sum_{i=1}^n \mathbb{E}[\eta(X_i)] \leq L$, then we may formulate a communication problem of communicating a message point over a memoryless channel with causal feedback. First note that the mutual information between the message point and observations is given by

$$\begin{aligned} \frac{1}{n} I(W; Y^n) &= \frac{1}{n} \sum_{i=1}^n I(W; Y_i | Y^{i-1}) \\ &= \sum_{i=1}^n \mathbb{E} \left[\log \frac{dB_{i|i}}{dB_{i|i-1}}(W) \right] \end{aligned}$$

Shannon’s converse to the channel coding theorem with feedback tells us that in order to achieve capacity, this aforementioned quantity must asymptotically be maximized. This allows for us to consider the following maximization problem

$$\max_{\pi \in \Pi} I(W; Y^n) + \alpha \mathbb{E}_\pi \left[\sum_{i=1}^n \eta(X_i) \right],$$

where α serves as a Lagrange multiplier such that under an optimal policy, the average state cost is upper bounded by L . We note that this can be captured in a causal coding/decoding framework by considering the sequential information gain distortion function (34). From Lemma VI.10, we note that a sufficient condition for optimality to this control problem is for

- $I(Y_i; Y^{i-1}) = 0$ for all i
- $X_i \sim P_X^*(\eta, P_{Y|X}, L)$, given in (11), for all i

Let $X = \mathbb{R}$ and denote $F_X(\cdot)$ as the cumulative distribution function of the optimal input distribution $P_X^*(\eta, P_{Y|X}, L)$. The Posterior matching (PM) scheme [15] simultaneously enables the two properties to hold for each i and is given by:

$$X_i = F_X^{-1}(F_{W|Y^{i-1}}(W|Y^{i-1})) \quad (74a)$$

$$= F_X^{-1}(B_{i-1|i-1}([0, W])) \quad (74b)$$

$$= \bar{e}(W, Z_{i-1}) \quad (74c)$$

where (74c) follows from Theorem V.3 and because $W_i = W$. Note that the $F_{W|Y^{i-1}}$ operation constructs a uniform-[0, 1] random variable that is independent of the past channel outputs, and the F_X^{-1} shaping operation enables each input to be drawn according to the optimal channel input distribution $P_X^*(\eta, P_{Y|X}, L)$. Note that from (74c), the PM scheme can be interpreted as ‘minimal’ from our structural result in the causal coding/decoding framework in Section V. Moreover, the causal encoder is *time-invariant*, and so likewise for the decoder acting as the nonlinear filter; thus, this means the PM scheme also can be interpreted as an instance of the inverse optimal control framework via Lemma VI.10. See Figure 9 and its relationship with Figure 5. Also note that in

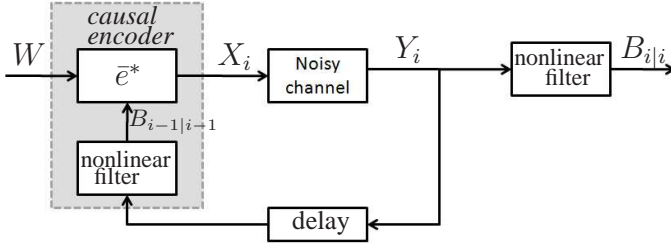


Fig. 9. Posterior matching scheme by Shayevitz & Feder, interpreted as a time-invariant manifestation of the simplified structural result in Figure 5.

some cases, the time-reversibility condition for inverse optimal control in Section VI-A is applicable: from Example 1, the PM scheme (74) elicits reversibly feasible dynamics. Consider the additive Gaussian noise channel. Under the PM scheme, $(X_i, Y_i : i \geq 1)$ are jointly Gaussian (see (53e) and [13, Example 1]). Note that since $X_i = \tilde{X}_{i-1}$ in (53d), the joint reversibility sufficient condition is equivalent to reversibility of the Markov chain X . Since all stationary Gaussian processes are time-reversible, we see that in this scenario, the time-reversibility framework for our inverse optimal control framework is linked to the PM scheme. Although our control problem only addresses the maximization of mutual information - which is a necessary condition for reliable communication by the converse to the channel coding theorem - it can be shown that reliable communication, as defined in Section I-A, results as a consequence of the mutual information maximization control problem under mild technical conditions [15].

C. Structural Result: Brain-Machine Interfaces

A brain-machine interface (BMI) is a system that elicits a direct communication pathway between a human and an external device. In many cases, it is the objective of the human to control an external device merely by imagination, and the external device acquires neural signals, actuates some

physical system, and perceptual feedback is given to the user to complete the loop. We now demonstrate how our structural result can be applied to the design of brain-machine interfaces that have a ‘user-friendly’ structure: displaying the minimal amount of useful perceptual feedback to the user, and designing an interaction strategy between the user and the external device.

Consider a brain-machine interface where a human has a desired high-level intent represented by the Markov process $(W_i : i \geq 1)$. At each time step, the human imagines a control signal X_i which is statistically linked to neural activity Y_i that is observed by the external device. For example, the statistics of Y_i are different when imagining a left-oriented movement $X_i = 0$ as compared to imagining a right-oriented movement $X_i = 1$ [70]. At each time step, the external device maps all its recorded observations Y^i to actuate some system, whose state is given by Z_i . Equally as important, the user gets perceptual feedback from the external device and allows this, along with causal information about the high-level intent, W^i , to specify the subsequent imagined control signal X_i .

Without loss of generality, because we do not know yet what perceptual feedback is the most relevant, we could consider a scenario where all information available to the decoder at any time i is fed back to the subject. Secondly, we may assume that we are planning to design the coordination strategy between the user and the interface: not only how the interface should take its observations and actuate the plant, but also what perceptual feedback should be specified back to the user *and* how the user should react to the perceptual feedback to specify the subsequent control signal X_i . In such a case, this problem boils down to our problem formulation in Section III. Note that because of the causal nature of the problem, real-time constraints with a human in the loop obviate the possibility of using ‘block-coding’ like paradigms. Secondly, such settings are more complicated than simply optimally representing intent with an arithmetic coding procedure as in [71] - because of the inherent uncertainty also due to the noisy channel mapping intent to neural signals.

Almost all previous approaches to design BMIs failed to consider how the desired control signals change in response to sensory feedback. For example, many previous schemes simply attempt to recursively estimate X_i from Y^i under the assumption that $(X_i : i \geq 1)$ is a Markov process. However, as we know from our structural result, for an arbitrary objective with additive cost function, it is crucially important for the system to keep a running estimate, or belief, on W_i given Y^i . Moreover, it is critically important that the user and the system agree on an interaction protocol that specifies both *what* sensory feedback is provided to the user (e.g. the sufficient statistics) and *how* the user should react to this feedback in pursuit of high-level intent (e.g. the function \bar{e}_i).

Our structural result says that first a state filter can construct sufficient statistics $S_i = (Z_{i-1}, B_{i|i})$, and then the external device can actuate the plant using S_i and the user only needs to be fed back S_{i-1} as perceptual feedback. This information, along with the current high-level goal Z_i , is all that is needed to specify an optimal causal encoder \bar{e}_i . See Figure 10.

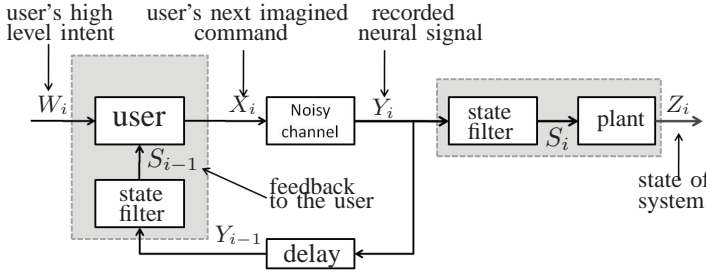


Fig. 10. Structural result within the context of a brain-machine interface: in an optimal system, the user acts as part of the causal encoder. The other part accumulates all causal observations and summarizes them into sufficient statistics acting as perceptual feedback to the user.

In [53], we instantiate this idea in an EEG-based BMI in two steps. We assume the high-level intent can be mathematically represented as a Markov process ($W_i = W : i \geq 1$) on $W = [0, 1]$ for which W is uniformly distributed over the $[0, 1]$ line. As such, this means we are assuming that the whole high-level intent is known to the user at all times. To relate this to a variety of practical applications, the user interprets the message point as a countably infinite sequence of symbols $D = (D_1, D_2, \dots)$ in an *ordered* countable set \mathcal{D} with a known statistical model (typically modeled as a fixed-order Markov process). Examples of the sequence D include an infinite sequence of text characters or an infinite sequence of small path arcs pertaining to a smooth path of bounded curvature. We use arithmetic coding [2] to develop a one-to-one mapping between any such sequence D and a point $W = \tau(D)$ uniformly distributed on the $[0, 1]$ line. We subsequently use an EEG system and specify a binary-input (left/right motor imagery) noisy channel with a spatial filter to extract beliefs $B_{i|i}$ sequentially [70]. With this, we implement the Posterior Matching scheme for the binary symmetric channel [16], [15]. Here, what is nice for a human in the loop is that $\bar{e}_i = \bar{e}$, and secondly, for the BSC, it only requires a *functional* of the posterior $B_{i-1|i-1}$ to be given to the encoder at time i : the median (denoted as $m(B_{i-1|i-1})$) [16], [13]:

$$X_i = \begin{cases} 0, & W < m(B_{i-1|i-1}) \\ 1, & W \geq m(B_{i-1|i-1}) \end{cases} \quad (75)$$

Because of the one-to-one mapping τ , at time i , this can be implemented by visually displaying the median path $\tau^{-1}(m(B_{i-1|i-1}))$ on the screen and instructing the user to obey the time-invariant PM scheme (75) within the context of the median path. This simply means performing a lexicographically comparison to D (i.e. identify the first location where the sequences differ and perform a symbol-based comparison). We have successfully implemented this to demonstrate reliable text spelling and two-dimensional smooth path specification. Secondly, wedding with arithmetic coding with the PM scheme has the added benefit that a natural ‘propagation’ of uncertainty ensues: the locations where D and $\tau^{-1}(m(B_{i-1|i-1}))$ differ increase to later and later parts of their sequences; this leads to a natural real-time implementation plausibility. Remote-control of an unmanned aerial vehicle using this paradigm has recently been shown in [72].

We also comment how the PM scheme by Shayevitz and Feder [15] is particularly relevant here: formulating this prob-

lem as one where the encoder has one of 2^{nR} hypotheses would mean that the human agent attempting to elicit neural control of an external device would have to implement an a strategy that differentiates possible inputs based upon one of 2^{nR} hypotheses. Even with visualization, this could be cumbersome. Moreover, it is unclear how the design specification would change when $n = 100$ as compared to when $n = 101$. Remarkably, using the posterior matching framework makes this problem truly solvable both theoretically and practically - by simply changing the starting point to be $W = [0, 1]$ and $Z = \mathcal{P}(W)$ and defining an appropriate information gain cost criterion. These observations speak to the fragility at which information theoretic problems with the same fundamental limits can be formulated.

The structural result demonstrated in this paper now enables the opportunity to design many brain-machine interface paradigms for a variety of cost functions beyond the the information gain paradigm and with assumption that $W_i = W_{i-1}$. The structural result has the potential more generally to enable an interesting intersection of desires on one platform: (i) guaranteed optimality from a decision-theoretic viewpoint; (ii) elucidation of the minimal amount of perceptual feedback information required to optimally display to the user; and (iii) potential ease-of-use when (e.g. when $\bar{e}_i = \bar{e}$ and it has a simple operational interpretation).

D. Inverse Optimal Control: Gauss-Markov source and AGN channel

Here we show that a stationary Markov coordination strategy consisting of a linear ‘estimation error’ encoder and MMSE decoder is inverse-control optimal for a Gauss-Markov Q_W and a power-constrained additive Gaussian channel. A variant of this problem for $\rho(w_i, z_{i-1}, z_i) \equiv \rho(w_i, z_i) = (w_i - z_i)^2$ has been studied by [39],[20].

Let $W = X = Y = Z = \mathbb{R}$. The source is a Gauss-Markov process with i.i.d. $\tilde{W}_i \sim \mathcal{N}(0, \sigma_m^2)$,

$$W_0 \sim \mathcal{N}\left(0, \frac{\sigma_m^2 \sigma_v^2}{L + \sigma_v^2(1 - \rho^2)}\right), \quad (76a)$$

$$W_i = \rho W_{i-1} + \tilde{W}_i \quad i \geq 1, \quad (76b)$$

$$I(\tilde{W}_i; X^{i-1}, Y^{i-1}) = 0, \quad i \geq 1 \quad (76c)$$

Note that we are not assuming that W is stationary. As such, this problem can be connected to problems in ‘control over noisy channels. In such problems with quadratic cost and linear Gaussian dynamics, the essence of optimally solving this control over noisy channels problem is optimally solving this causal coding/decoding ‘active tracking’ problem [39],[20].

The channel additive with Gaussian noise (AGN):

$$Y_i = X_i + V_i \quad V_i \sim \mathcal{N}(0, \sigma_v^2) \quad (77)$$

A typical objective in practice is to design an encoder and decoder than can minimize the mean-squared error in estimating the source process, i.e., minimize $J(e^n, d^n) = \mathbb{E} \left[\sum_{i=1}^n (Z_i - W_i)^2 + \alpha X_i^2 \right]$. It is known [39],[20] that an

optimal linear coordination strategy exists, pertaining to “error” encoding and MMSE estimation decoding:

$$X_i = \beta_i (W_i - \mathbb{E}[W_i|Y^{i-1}]) \quad (78a)$$

$$Z_i = \mathbb{E}[W_i|Y^i] \quad (78b)$$

where β_i are time-varying normalizing constants that result in $X_i \sim \mathcal{N}(0, L)$ for all i , and the power-constraint L depends on the value of α .

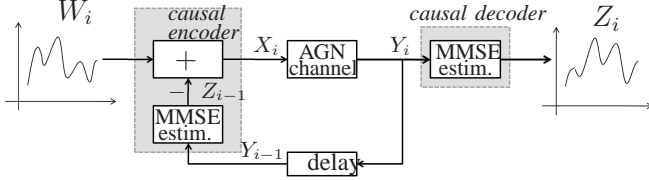


Fig. 11. With Q_W Gauss-Markov and $P_{Y|X}$ an AGN channel, “error” encoding and MMSE estimation decoding is inverse control optimal. The induced cost function is squared error.

We now consider observing this problem from the lens of inverse optimal control for a distortion function of the form $\rho(w_i, z_{i-1}, z_i)$:

Lemma VII.2. *For the problem setup in (76), define the following Stationary Markov coordination policy*

$$X_i = \beta (W_i - \rho Z_{i-1}) \quad (79a)$$

$$Z_i = \rho Z_{i-1} + \gamma Y_i \quad (79b)$$

where $\beta = \sqrt{\frac{L}{C}}$, $\gamma = \frac{\sqrt{LC}}{L + \sigma_v^2}$, and $C = \frac{\sigma_m^2}{1 - \rho^2 \frac{\sigma_v^2}{L + \sigma_v^2}}$.

- (a) The policy pair in (79) is inverse control optimal

$$\eta(x_i) \propto_+ x_i^2 \quad (80a)$$

$$\rho(w_i, z_{i-1}, z_i) \propto_+ (w_i - z_i)^2 - \frac{\sigma_v^2}{L + \sigma_v^2} (w_i - z_{i-1})^2 \quad (80b)$$

- (b) The total cost can be represented as a weighted MMSE cost given by $J_{n,\pi}^\alpha$:

$$\begin{aligned} & \mathbb{E} \left[\sum_{i=1}^n \rho(W_i, Z_{i-1}, Z_i) \right] \\ &= \mathbb{E} \left[\sum_{i=1}^n (Z_i - W_i)^2 + \left(\frac{1}{1 - \frac{\sigma_v^2 \rho^2}{L + \sigma_v^2}} \right) (Z_n - W_n)^2 \right] \end{aligned}$$

The proof is provided in Appendix H.

Remark 5. *The policy-pair in (78) is optimal for a mean-square distortion cost (MMSE) problem for Gauss-Markov sources except that the last reconstruction has higher penalty. For $n \rightarrow \infty$, the cost for which (78) is optimal is exactly equivalent to a MMSE cost problem $\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E} \left[\sum_{i=1}^n (Z_i - W_i)^2 + \alpha X_i^2 \right]$. Thus, asymptotically, we can recover the results of [39], [20, Ch. 6] using inverse-optimal control and time-invariant cost functions.*

E. Inverse Optimal Control: the M/M/1 Queue

Here we show that the $\cdot/M/1$ queue’s dynamics can be interpreted as a stationary Markov coordination strategy that is inverse control optimal for Q_W being a Poisson process. It is well-known from Burke’s theorem [64], [67] that for a Poisson process of rate λ entering a $\cdot/M/1$ queue, in steady state the queue state at time t is independent of the output before time t . We now demonstrate that this statement has implications not only for the capacity of queuing timing channels [54], [55], [68], [73], but also for inverse optimal control.

Divide time into units of interval Δ where $\Delta \ll 1$. The input to the queue W_i represents the number of arrivals to the queue till time i . For a Poisson source, $(W_i : i \geq 1)$ is the discrete-time equivalent of the counting function representation of a Poisson process.

$$Q_W(w_i | w_{i-1}) = \begin{cases} \lambda \Delta, & \text{if } w_i = w_{i-1} + 1 \\ 1 - \lambda \Delta, & \text{if } w_i = w_{i-1} \\ 0, & \text{otherwise} \end{cases} \quad (81)$$

In other words, $W_i = W_{i-1} + \tilde{W}_i$ where \tilde{W}_i are i.i.d. with $\mathbb{P}(\tilde{W}_i = 1) = \lambda \Delta$. Assume the following model for the channel:

$$P_{Y|X}(1|x) = \begin{cases} 0 & x = 0 \\ \mu \Delta & x > 0 \end{cases} \quad (82)$$

For a queuing system, note that this means that a departure ($Y_i = 1$) can only occur when the number of customers in the queue is positive, and the likelihood of a departure in that scenario for a bin of length Δ is $\mu \Delta$. Continuing on with the queuing analogy, note that we represent Z as the counting function representation of the departure process as $Z_i = \sum_{k \leq i} Y_k$ where $Y_k \in \{0, 1\}$. X_i is the queue size representing the number of customers in the queue at the i -th time instant: $X_i = W_i - Z_{i-1}$. Thus, the update equations for the state X_i and output of the queue Z_i are linear stationary Markov policies given by

$$X_i = W_i - Z_{i-1} \quad (83a)$$

$$Z_i = Z_{i-1} + Y_i \quad (83b)$$

The departure at i -th time instant Y_i depends on the state by the following discrete memoryless ‘Z’ channel model: That is, there will be no departure if the queue is empty, and there will be departure with probability $\mu \Delta$ if the queue is not empty. The initial number of arrivals W_0 is drawn according to $\mathbb{P}(W_0 = k) = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^k$, $k \geq 0$ and the initial number of departures $Z_0 = 0$. Note that the aggregate statistical dynamics of $P_{Z^n|W^n}^{\pi}$ in Figure 12 are precisely that of the discrete-time exponential server timing channel, also termed a $\cdot/M/1$ queue of rate μ , which is a first-come, first-serve queuing system with i.i.d. service times geometrically distributed of rate μ [64]. As $\Delta \rightarrow 0$, this becomes the continuous-time $\cdot/M/1$ queue. From standard queuing theory it follows that X is a birth-death Markov chain in steady-state with distribution

$$\pi_k = \left(1 - \frac{\lambda}{\mu}\right) \left(\frac{\lambda}{\mu}\right)^k, \quad k \geq 0. \quad (84)$$

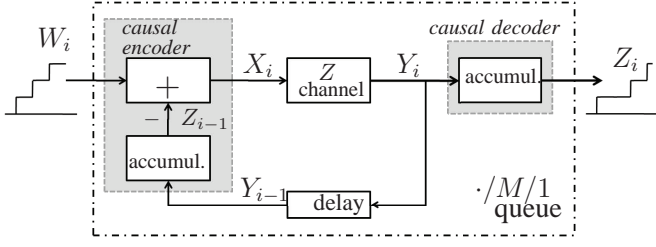


Fig. 12. With Q_W a Poisson process and $P_{Y|X}$ a Z channel, the $\cdot/M/1$ queue is inverse control optimal.

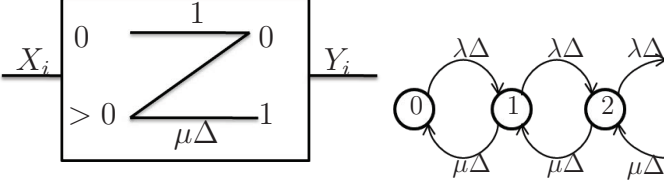


Fig. 13. $P_{Y|X}$ for the $\cdot/M/1$ queue sampled at length- Δ intervals

Fig. 14. birth-death chain for X in the $M/M/1$ queue

Therefore Lemma VI.8 holds and note that the fixed coordination strategy given by (83) is inverse-control optimal for

$$\rho(w_i, z_{i-1}, z_i) = \begin{cases} -\log(1 - \lambda\Delta), & x_i = 0, y_i = 0; \\ \log \frac{1-\mu\Delta}{1-\lambda\Delta}, & x_i > 0, y_i = 0; \\ \log \frac{\mu}{\lambda}, & x_i > 0, y_i = 1; \\ +\infty, & \text{otherwise.} \end{cases}$$

$$\Rightarrow \lim_{\Delta \rightarrow 0} \rho(w_i, z_{i-1}, z_i) = \begin{cases} 0, & y_i = 0 \\ \log \frac{\mu}{\lambda}, & x_i > 0, y_i = 1 \\ \infty & \text{otherwise} \end{cases} \quad (85)$$

Figure 12 is akin to [54, Fig 4], where it is shown that this insight and (85) leads to the derivation of the capacity of the exponential server timing channel.

Remark 6. Though the ESTC is time-varying, non-memoryless, and has non-linear dynamics from a inter-arrival time viewpoint [54], when viewed appropriately, its internal structure consists of a time-invariant memoryless ‘Z’ channel and a feedback loop comprising a linear SM coordination strategy $\bar{\pi}$. Moreover, for a Poisson process input, $\bar{\pi}$ is inverse control optimal. As such, the internal structure of the $\cdot/M/1$ queue can be interpreted as an optimal decentralized controller. Also, note how the internal structure is exactly synonymous to the Gaussian case (76) ([39],[20]) in that the encoder and decoder are both linear dynamical systems.

The result differs from the source-channel matching results in [49, Sec 3] for two reasons: i) the problem is approached through an inter-arrival viewpoint in [49], while we use counting function representation (inputs and outputs to the queue). ii) The dynamics of the ESTC are fixed and [49] considers a possible encoder between the poisson process and the ESTC input, and a decoder between ESTC output and the reconstruction and show that the encoder and decoder should be identity mappings. In our case, the linear encoder and decoder policies are fixed and internal to the structure of the queue dynamics. As a consequence, the source-channel

matching results has to be performed over a less complicated memoryless ‘Z’ channel.

Other extensions to queuing timing channels fit within this framework as well: see for example the variety of queuing systems in [69] for which joint reversibility holds. Similar results hold for other queuing timing channels, such as:

- $\cdot/M/c$ queue: There are c servers each with an i.i.d exponential service time. In this case, the queue dynamics-the linear encoder and the decoder will be the same (Fig 12). The structure of memoryless channel ($P_{Y|X}$) will depend on c .
- ‘The queue with feedback’ [64, p 204-205]. Here, with probability $1 - p_0$ departures from the queue instantaneously return to the input of the queue (independent of all other processes). The ‘effective’ Z channel changes $\mu\Delta$ to $p_0\mu\Delta$ and all other arguments hold.

F. Inverse Optimal Control: Blackwell’s Trapdoor Channel

Here we show that the internal structure of Blackwell’s trapdoor channel can be interpreted as a stationary Markov coordination strategy that is inverse control optimal.

Consider ‘the chemical (trapdoor) channel’ [56], [57], [58] as shown in Fig 15. Initially (Fig. 15a), a ball labeled either 0 (red) or 1 (blue) is present in one of the two slots. Then (Fig. 15b) a ball, either a 0 or 1, is placed in the empty slot, after which (Fig. 15c) one of the trapdoors opens at random with probability $(\frac{1}{2}, \frac{1}{2})$. The ball lying above the open door then falls through. The door closes (as in Fig. 15a) and the process is repeated.

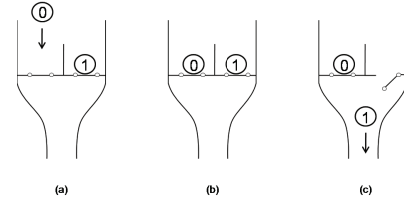


Fig. 15. Blackwell’s Trapdoor Channel

Let $\tilde{W}_i \in \{0, 1\}$ and $Y_i \in \{0, 1\}$ represent the color of the ball that is input and output of the trapdoor respectively. Define the channel input X_i to pertain to the composition of balls before one of the doors is opened (Fig 15b). That is, $X_i \in \{0, 1, 2\}$ where $X_i = 0$ represents two red balls (0,0), $X_i = 1$ represents a blue ball and a red ball (0,1) and $X_i = 2$ represent two blue balls (1,1). Thus, the dynamics are given by

$$X_i = X_{i-1} + \tilde{W}_i - Y_{i-1} \quad (86)$$

From a counting function viewpoint, let $\{W_i\}$ and $\{Z_i\}$ be the counting processes representing the number of blue balls that were input and output from the system. Hence X_i , as defined above tells about the composition of balls, or equivalently the

number of blue balls that are ‘in’ the system at time i .

$$W_i = W_{i-1} + \tilde{W}_i \quad (87a)$$

$$Z_i = Z_{i-1} + Y_i \quad (87b)$$

$$\begin{aligned} X_i &= X_{i-1} + \tilde{W}_i - Y_{i-1} \\ &= W_i - Z_{i-1} \end{aligned} \quad (87c)$$

Note that the state-update equation and the decoding policy (87b)-(87c) are reversibly feasible dynamics by Example 2. The output depends on the state according to the channel law $P_{Y|X}(Y|X)$ as (the inverse erasure channel) as shown in figure 17:

$$P_{Y|X}(1|x) = \begin{cases} 0 & x = 0 \\ \frac{1}{2} & x = 1 \\ 1 & x = 2 \end{cases} \quad (88)$$

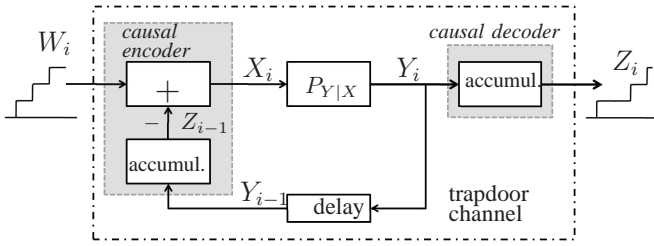


Fig. 16. With Q_W a Markov counting process (i.i.d. \tilde{W}_i inputs) and an ‘inverted E’ channel, Blackwell’s trapdoor channel is inverse control optimal.

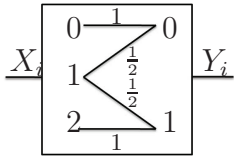


Fig. 17. $P_{Y|X}$ for the trapdoor channel

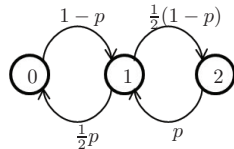


Fig. 18. birth-death chain for X in the trapdoor channel with Q_W a Markov counting process.

Fixing \tilde{W} to be an i.i.d process, with $\mathbb{P}(\tilde{W}_i = 0) = p$, and $Z_0 = 0$. The transition probabilities of the Markov Chain X_i are given by Fig 18

$$P = \begin{bmatrix} p & 1-p & 0 \\ \frac{1}{2}p & \frac{1}{2} & \frac{1}{2}(1-p) \\ 0 & p & 1-p \end{bmatrix}$$

and if W_0 is drawn according to

$$\mathbb{P}(W_0 = k) = \begin{cases} p^2, & k=0; \\ 2p(1-p), & k=1; \\ (1-p)^2, & k=2; \\ 0, & \text{otherwise.} \end{cases},$$

it follows that we have a birth-death chain initially in steady-state with distribution $\pi(\cdot) = \mathbb{P}(W_0 = \cdot)$. Thus from Lemma VI.8, we have that $\tilde{\pi}$ is inverse-control optimal. Moreover, the from Corollary VI.6, the trapdoor policy (86)

is optimal for the cost function of the form

$$\rho(w_i, z_{i-1}, z_i) = \begin{cases} \log p, & x_i = 0, y_i = 0; \\ \log 2p, & x_i = 1, y_i = 0; \\ \log 2(1-p), & x_i = 1, y_i = 1; \\ \log(1-p), & x_i = 2, y_i = 1; \\ +\infty, & \text{otherwise.} \end{cases}$$

Note that when $p = \frac{1}{2}$, $\mathbb{E}[\rho(W, Z_{i-1}, Z_i)] = -I(\pi, P_{Y|X}) = -\frac{1}{2}$, which coincides with the achievable rate coding scheme developed for the trapdoor channel in [57].

VIII. Discussion and Conclusion

In this paper, we have developed a new class of causal coding/decoding problems that can be understood from the lens of both information theory as well as control theory. We would like to emphasize that the primary focus of this paper is not about fundamental limits (although some new fundamental limits are presented). Rather, it is about attempting to develop a modeling framework whereby principles of traditional information theory (e.g. KL divergence, mutual information bounds, etc) and traditional concepts (e.g. dynamic programming, structural results), can be wed to elucidate things and impact the design of future real-world systems and applications.

A second aim is to demonstrate that by first formulating problems in this manner (whereby the notion of rate is not necessarily embedded directly in the problem formulation), fundamental information theoretic limits fall out as a consequence of solving the problem. For example, in our information gain cost, we demonstrate how the posterior matching scheme by Shayevitz & Feder is an optimal solution to a causal coding/decoding problem. We did *not* attempt to directly impose the notion of achievability in the cost function, but rather constructed a cost function from the converse. It was shown in [15] that in essence achievability still holds. Moving forward, this suggests that perhaps equally as much attention should be paid to the manner in which problems are formulated as what is being paid to attempt to solve already-formulated and un-solved problems. The authors believe that a significant amount of practical and theoretical advances - including and extending beyond communication - can be made if the information theory community embraces this challenge.

In light of how the second law of thermo-dynamics appeared in Section V and how Markov chain time-reversibility appeared in Section VI-A, perhaps further work could be pursued to further understand the relationship between information theory and thermodynamics. It has recently been suggested that such an understanding could additionally play a role in understanding brain function [74], [75]. Recent developments in the neuroscience community have begun to posit that Bayesian decision-making could implicitly be playing a role in the processes of sensation, perception, and decision-making in the mammalian brain through interacting neural systems sequentially handing to one another ‘what is missing’ [76][77][78], [79]. The sequential information gain framework perhaps could provide insight into further considering these matters. Analogously, the inverse optimal control framework

developed here, when combined with statistically inferring the coordination policies as in [43], [44], [45], [46], [6], [7], [8], could provide insight into what cost is being minimized.

IX. ACKNOWLEDGEMENTS

The authors thank T. Basar, T. Bretl, B. Hajek, N. Kiyavash, P. R. Kumar, A. Mahajan, P. Mehta, S. Meyn, M. Raginsky, R. Srikant, and S. Yüksel for useful discussion.

APPENDIX A PROOF OF LEMMA IV.1

Proof: As described in the statement of the lemma, define the state space $\mathbf{S} = \mathbf{Z} \times \mathcal{P}(\mathbf{W})$ and control space $\mathbf{U} = \tilde{\mathbf{E}} \times \mathbf{Z}$ with $s_i \in \mathbf{S}$, $u_i \in \mathbf{U}$ given by (27):

$$s_i = (z_{i-1}, b_{i|i}), \quad u = (\tilde{e}_{i+1}, z_i).$$

Then

$$\begin{aligned} & \mathbb{E} [\rho(W_i, Z_{i-1}, Z_i) | Z_{i-1} = z_{i-1}, Y^i = y^i, Z_i = z_i] \\ &= \int_{w_i \in \mathbf{W}} \rho(w_i, z_{i-1}, z_i) b_{i|i}(dw_i) \end{aligned} \quad (89)$$

$$\begin{aligned} & \equiv \tilde{\rho}(s_i, z_i) \\ & \mathbb{E} [\eta(X_{i+1}) | Z_{i-1} = z_{i-1}, Y^i = y^i, \tilde{E}_{i+1} = \tilde{e}_{i+1}] \\ &= \int_{w_{i+1} \in \mathbf{W}} \alpha \eta(\tilde{e}_{i+1}(w_{i+1})) b_{i+1|i}(dw_{i+1}) \\ &= \int_{w_{i+1} \in \mathbf{W}} \alpha \eta(\tilde{e}_{i+1}(w_{i+1})) \Phi(b_{i|i})(dw_{i+1}) \quad (90) \\ & \equiv \tilde{\eta}(s_i, \tilde{e}_{i+1}) \end{aligned}$$

where (89) follows from (19); (90) follows from (21). ■

APPENDIX B PROOF OF LEMMA IV.2

Proof: Note that

$$\begin{aligned} & P_{S_{i+1} | S^i = s^i, U^i = u^i}(ds_{i+1}) \\ &= 1_{\{s_{i+1,1} = u_{i,2}\}} \int_{w_{i+1} \in \mathbf{W}} \int_{y_{i+1} \in \mathbf{Y}} 1_{\{b_{i+1|i+1} = \Lambda(b_{i|i}, y_{i+1}, \tilde{e}_{i+1})\}} \\ & \quad P_{Y|X}(dy_{i+1} | \tilde{e}_{i+1}(w_{i+1})) b_{i+1|i}(dw_{i+1}) \quad (91a) \end{aligned}$$

$$\begin{aligned} &= 1_{\{s_{i+1,1} = u_{i,2}\}} \int_{w_{i+1} \in \mathbf{W}} \int_{y_{i+1} \in \mathbf{Y}} 1_{\{s_{i+1,2} = \Lambda(s_{i,2}, y_{i+1}, u_{i,1})\}} \\ & \quad P_{Y|X}(dy_{i+1} | u_{i,1}(w_{i+1})) \Phi(s_{i,2})(dw_{i+1}) \quad (91b) \end{aligned}$$

$$\begin{aligned} &= P_{S_{i+1} | S_i = s_i, U_i = u_i}(ds_{i+1}) \\ &\equiv Q_S(ds_{i+1} | s_i, u_i) \quad (91c) \end{aligned}$$

where (91a) follows from (17) and (14); (91b) follows from (27); and (91c) demonstrates that this is a controlled Markov

chain with *time-invariant* statistical dynamics. ■

APPENDIX C PROOF OF LEMMA V.1

Proof:

$$I(W^n; Y^n) = \sum_{i=1}^n I(W^n; Y_i | Y^{i-1}) \quad (92)$$

$$\begin{aligned} &= \sum_{i=1}^n I(W^i; Y_i | Y^{i-1}) \\ & \quad + I(W_{i+1}^n; Y_i | Y^{i-1}, W^i) \quad (93) \end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^n I(W^i; Y_i | Y^{i-1}) \\ & \quad + I(W_{i+1}^n; Y_i | Y^{i-1}, W^i, X_i) \quad (94) \end{aligned}$$

$$= \sum_{i=1}^n I(W^i; Y_i | Y^{i-1}) \quad (95)$$

$$\begin{aligned} &= \sum_{i=1}^n I(W_i; Y_i | Y^{i-1}) \\ & \quad + I(W^{i-1}; Y_i | W_i, Y^{i-1}) \quad (96) \end{aligned}$$

$$\begin{aligned} &= \sum_{i=1}^n I(W_i; Y_i | Y^{i-1}) \\ & \quad + I(W^{i-1}; Y_i | W_i, Y^{i-1}, X_i) \quad (97) \end{aligned}$$

$$= \sum_{i=1}^n I(W_i; Y_i | Y^{i-1}) \quad (98)$$

where (92) follows from (10); (93) follows from (10); (94) follows from (15); (95) follows from (17); (96) follows from (10); (97) follows from our assumption (32a) that the encoder operates on sufficient statistics; and (98) follows from (17). ■

APPENDIX D PROOF OF LEMMA V.2

Proof: (35) follows directly from Lemma IV.1. Now, let us focus on (36). From Lemma IV.1, we have that

$$\begin{aligned} \tilde{\rho}(s, z) &= \int_{w \in \mathbf{W}} \rho(w, b, z) b'(dw) \\ &= \int_{w \in \mathbf{W}} -\log \frac{dz}{d\Phi(b)}(w) b'(dw) \quad (99) \end{aligned}$$

where (99) follows from (34) for any z satisfying $z \ll \Phi(b)$ and is infinite otherwise. Now note that if it is not the case that $b' \ll z$, then there exists a set $A \in \mathcal{B}(\mathbf{W})$ for which $z(A) = 0$ and $b'(A) > 0$ and thus it follows that $\frac{dz}{d\Phi(b)}(w) = 0 \Rightarrow -\log \frac{dz}{d\Phi(b)}(w) = \infty$ for all $w \in A$. Thus $\tilde{\rho}(s, z) = \infty$. Now assume $b' \ll z \ll \Phi(b)$. Then since if $\beta \ll \nu \ll \mu$ then $\frac{d\beta}{d\mu} = \frac{d\beta}{d\nu} \frac{d\nu}{d\mu}$, μ -almost everywhere [80, Sec 5.5], it follows that

$$\begin{aligned} \tilde{\rho}(s, z) &= \int_{w \in \mathbf{W}} -\log \frac{db'}{d\Phi(b)}(w) b'(dw) + \log \frac{db'}{dz}(w) b'(dw) \\ &= -D(b' \| \Phi(b)) + D(b' \| z) \quad (100) \end{aligned}$$

■

Proof: In order to find the optimal cost J_n^* given by $J_n^* = \mathbb{E}[V_0(S_0)]$, we use the standard dynamic programming approach and evaluate optimal cost-to-go functions $\{V_k : k = 0, \dots, n\}$. Consider the final-stage problem of finding $V_n(s_n)$, where $s_n = (z_{n-1}, b_{n|n})$ and describe any control u_n as $u_n = (\tilde{e}_{n+1}, z_n)$. Then the one-stage problem is

$$\begin{aligned} V_n(s_n) &= \inf_{u_n=(\tilde{e}_{n+1}, z_n)} \bar{g}_n(s_n, u_n) \\ &= \inf_{z_n \in \mathcal{P}(\mathcal{W})} \bar{\rho}(s_n, z_n) \end{aligned} \quad (101)$$

$$= -D(b_{n|n} \parallel \Phi(z_{n-1})) + \inf_{z_n \in \mathcal{P}(\mathcal{W}), b_{n|n} \ll z_n \ll \Phi(z_{n-1})} D(b_{n|n} \parallel z_n) \quad (102)$$

$$= -D(b_{n|n} \parallel \Phi(z_{n-1})) \quad (103)$$

where (101) follows (30); (102) follows from (36); and (103) follows from the non-negativity of the KL divergence. The optimal choice of z_n is the one for which the equality in (103) holds true and hence under an optimal policy, $z_n = b_{n|n}$. This follows the same reasoning that elicits how for in the self-information loss sequential probability assignment, the best probability assignment is the true belief [28].

For the second-step $k = n - 1$, consider finding $V_{n-1}(s_{n-1})$, where $s_{n-1} = (z_{n-2}, b_{n-1|n-1})$ and describe any control u_{n-1} as $u_{n-1} = (\tilde{e}_n, z_{n-1})$. Then we have:

$$V_{n-1}(s_{n-1}) = \inf_{u_{n-1}=(\tilde{e}_n, z_{n-1})} \bar{g}_{n-1}(s_{n-1}, u_{n-1}) + \mathbb{E}[V_n(z_{n-1}, B_{n|n}) | S_{n-1} = s_{n-1}, U_{n-1} = u_{n-1}] \quad (104)$$

$$= \inf_{\tilde{e}_n, z_{n-1}} \alpha \bar{\eta}(s_{n-1}, \tilde{e}_n) + \bar{\rho}(s_{n-1}, z_{n-1}) + \mathbb{E}[-D(B_{n|n} \parallel \Phi(z_{n-1})) | S_{n-1} = s_{n-1}, U_{n-1} = u_{n-1}] \quad (105)$$

$$\begin{aligned} &= -D(b_{n-1|n-1} \parallel \Phi(z_{n-2})) + \inf_{\tilde{e}_n} \alpha \bar{\eta}(s_{n-1}, \tilde{e}_n) + \inf_{b_{n-1|n-1} \ll z_{n-1} \ll \Phi(z_{n-2})} D(b_{n-1|n-1} \parallel z_{n-1}) \\ &\quad + \mathbb{E}[-D(B_{n|n} \parallel \Phi(z_{n-1})) | S_{n-1} = s_{n-1}, \tilde{E}_n = \tilde{e}_n] \end{aligned} \quad (106)$$

where (105) follows by substituting values of \bar{g}_{n-1} and V_n from (30) and (103); (106) follows from (36).

For any fixed encoder policy \tilde{e}_n , the optimal choice for z_{n-1} is to pick $z_{n-1} = b_{n-1|n-1}$ as shown:

$$z_{n-1}^*(s_{n-1}) = \arg \inf_{b_{n-1|n-1} \ll z_{n-1} \ll \Phi(z_{n-2})} D(b_{n-1|n-1} \parallel z_{n-1}) - \mathbb{E}[D(B_{n|n} \parallel \Phi(z_{n-1})) | S_{n-1} = s_{n-1}, \tilde{E}_n = \tilde{e}_n] \quad (107)$$

$$\begin{aligned} &= \arg \inf_{b_{n-1|n-1} \ll z_{n-1} \ll \Phi(z_{n-2})} D(P_\Lambda(\cdot | b_{n-1|n-1}, \tilde{e}_n) \parallel P_\Lambda(\cdot | z_{n-1}, \tilde{e}_n)) \\ &\quad + \underbrace{D(b_{n-1|n-1} \parallel z_{n-1}) - \mathbb{E}[D(\Lambda(b_{n-1|n-1}, Y_n, \tilde{e}_n) \parallel \Lambda(z_{n-1}, Y_n, \tilde{e}_n)) | S_{n-1} = s_{n-1}, \tilde{E}_n = \tilde{e}_n]}_{\geq 0} \end{aligned} \quad (108)$$

$$= b_{n-1|n-1}. \quad (109)$$

where (108) and the non-negativity of the difference follow because:

$$\begin{aligned}
z_{n-1}^*(s_{n-1}) &= \arg \inf_{b_{n-1|n-1} \ll z_{n-1} \ll \Phi(z_{n-2})} D(b_{n-1|n-1} \| z_{n-1}) - \mathbb{E} \left[D(B_{n|n} \| \Phi(z_{n-1})) \mid S_{n-1} = s_{n-1}, \tilde{E}_n = \tilde{e}_n \right] \\
&= \arg \inf_{b_{n-1|n-1} \ll z_{n-1} \ll \Phi(z_{n-2})} D(b_{n-1|n-1} \| z_{n-1}) \\
&\quad - \mathbb{E} \left[D(B_{n|n} \| \Phi(b_{n-1|n-1})) - \mathbb{E}_{B_{n|n}} \left[\log \frac{d\Phi(b_{n-1|n-1})}{d\Phi(z_{n-1})} \mid B_{n|n} \right] \mid B_{n-1|n-1} = b_{n-1|n-1} \right] \tag{110}
\end{aligned}$$

$$\begin{aligned}
&= \arg \inf_{b_{n-1|n-1} \ll z_{n-1} \ll \Phi(z_{n-2})} D(b_{n-1|n-1} \| z_{n-1}) \\
&\quad - \int_{y \in \mathcal{Y}} \int_{w \in \mathcal{W}} \log \frac{d\Phi(b_{n-1|n-1})}{d\Phi(z_{n-1})}(w) \underbrace{\Lambda(b_{n-1|n-1}, y, \tilde{e}_n)}_{b_{n|n}}(dw) P_\Lambda(dy | b_{n-1|n-1}, \tilde{e}_n) \tag{111}
\end{aligned}$$

$$\begin{aligned}
&= \arg \inf_{b_{n-1|n-1} \ll z_{n-1} \ll \Phi(z_{n-2})} D(b_{n-1|n-1} \| z_{n-1}) \\
&\quad - \int_{y \in \mathcal{Y}} \int_{w \in \mathcal{W}} \log \left(\frac{d\Lambda(b_{n-1|n-1}, y, \tilde{e}_n)}{d\Lambda(z_{n-1}, y, \tilde{e}_n)} \right) (w) \Lambda(b_{n-1|n-1}, y, \tilde{e}_n)(dw) P_\Lambda(dy | b_{n-1|n-1}, \tilde{e}_n) \\
&\quad + \int_{y \in \mathcal{Y}} \log \frac{dP_\Lambda(\cdot | b_{n-1|n-1}, \tilde{e}_n)}{dP_\Lambda(\cdot | z_{n-1}, \tilde{e}_n)}(y) P_\Lambda(dy | b_{n-1|n-1}, \tilde{e}_n) \tag{112}
\end{aligned}$$

$$\begin{aligned}
&= \arg \inf_{b_{n-1|n-1} \ll z_{n-1} \ll \Phi(z_{n-2})} D(P_\Lambda(\cdot | b_{n-1|n-1}, \tilde{e}_n) \| P_\Lambda(\cdot | z_{n-1}, \tilde{e}_n)) \\
&\quad + \underbrace{D(b_{n-1|n-1} \| z_{n-1}) - \mathbb{E}_{P_\Lambda(\cdot | b_{n-1|n-1}, \tilde{e}_n)} [D(\Lambda(b_{n-1|n-1}, Y_n, \tilde{e}_n) \| \Lambda(z_{n-1}, Y_n, \tilde{e}_n))]}_{\geq 0} \tag{113}
\end{aligned}$$

where (110) follows because $B_{n|n} \ll \Phi(b_{n-1|n-1}) \ll \Phi(z_{n-1})$ and so $\frac{dB_{n|n}}{d\Phi(z_{n-1})} = \frac{dB_{n|n}}{d\Phi(b_{n-1|n-1})} \frac{d\Phi(b_{n-1|n-1})}{d\Phi(z_{n-1})}$; (111) follows from the definition of the nonlinear filter (23b); (112) follows from the fact that $b_{n-1|n-1} \ll z_{n-1}$ and the definition of the nonlinear filter in (20); and the difference in (113) being non-negative follows from mapping this scenario to that of the hidden Markov model and the nonlinear filter:

- Here, the latent Markov process is W and one observation Y_n is recorded.
- Because in this dynamic programming problem, while in state s_{n-1} and under a fixed $\tilde{e}_n : \mathcal{W} \rightarrow \mathcal{X}$, the noisy channel from W_n to Y_n is the composition of the encoder map \tilde{e}_n and the input to the channel from X_n to Y_n : $P_{Y_n|W_n}(dy|w_n) = P_{Y|X}(dy|\tilde{e}_n(w_n))$.
- Two different decoders both know the statistical dynamics but have different initial beliefs about W_{n-1} . One decoder's initial belief is $b_{n-1|n-1} \in \mathcal{P}(\mathcal{W})$ and the other's is $z_{n-1} \in \mathcal{P}(\mathcal{W})$. The initial 'distance' between the beliefs is measured by the KL divergence, $D(b_{n-1|n-1} \| z_{n-1})$.
- Both decoders observe Y_n and update their beliefs about W_n according to the one-step nonlinear filter one updates its belief according to $\Lambda(b_{n-1|n-1}, Y_n, \tilde{e}_n)$ and the other does so according to $\Lambda(z_{n-1}, Y_n, \tilde{e}_n)$. The divergence between their beliefs after the observation is given by $D(\Lambda(b_{n-1|n-1}, Y_n, \tilde{e}_n) \| \Lambda(z_{n-1}, Y_n, \tilde{e}_n))$ and on average this is smaller than the original due to Jensen's inequality and the second law of thermodynamics for hidden Markov chains. This inequality is thus a manifestation of how the relative entropy is a 'Lyapunov function' for the stability (e.g. insensitivity to initial beliefs) of the nonlinear filter [50, Remark 4.2].

Hence the optimal choice for z_{n-1} is to pick $b_{n-1|n-1}$. Consequently,

$$\begin{aligned}
V_{n-1}(s_{n-1}) &= -D(b_{n-1|n-1} \| \Phi(z_{n-2})) + \inf_{\tilde{e}_n} \alpha \bar{\eta}(s_{n-1}, \tilde{e}_n) + \mathbb{E} \left[V_n(b_{n-1|n-1}, B_{n|n}) \mid B_{n-1|n-1} = b_{n-1|n-1}, \tilde{E}_n = \tilde{e}_n \right] \\
&= -D(b_{n-1|n-1} \| \Phi(z_{n-2})) + \alpha \bar{\eta}(s_{n-1}, \tilde{e}_n^*[b_{n-1|n-1}]) \\
&\quad + \mathbb{E} \left[V_n(b_{n-1|n-1}, B_{n|n}) \mid B_{n-1|n-1} = b_{n-1|n-1}, \tilde{E}_n = \tilde{e}_n^*[b_{n-1|n-1}] \right] \tag{114}
\end{aligned}$$

Using an inductive argument and the exact same set of arguments as above, it follows that for any $1 \leq k \leq n-1$, and any encoder policy \tilde{e}_{k+1} , the optimal choice for z_k is given by $b_{k|k}$ and that for $s_k = (z_{k-1}, b_{k|k})$,

$$\begin{aligned}
V_k(s_k) &= -D(b_{k|k} \| \Phi(z_{k-1})) + \alpha \bar{\eta}(s_k, \tilde{e}_{k+1}^*[b_{k|k}]) \\
&\quad + \mathbb{E} \left[V_{k+1}((b_{k|k}, B_{k+1|k+1})) \mid B_{k|k} = b_{k|k}, \tilde{E}_{k+1} = \tilde{e}_{k+1}^*[b_{k|k}] \right] \tag{115}
\end{aligned}$$

For the initial step, $k = 0$, by definition $Z_0(A) = B_{0|0}(A)\mathbb{P}(W_0 \in A)$ and is known to both encoder and decoder. Thus the

minimization is only over \tilde{e}_1 . For a state $s_0 = (z_{-1}, b_{0|0})$ and control $u_0 = (\tilde{e}_1, z_0) = (\tilde{e}_1, b_{0|0})$, we have:

$$V_0(s_0) = \inf_{\tilde{e}_1} \alpha \bar{\eta}(s_0, \tilde{e}_1) + \mathbb{E} \left[V_1(b_{0|0}, B_{1|1}) | B_{0|0} = b_{0|0}, \tilde{E}_1 = \tilde{e}_1 \right] \quad (116)$$

$$= \alpha \bar{\eta}(s_0, \tilde{e}_1^*[b_{0|0}]) + \mathbb{E} \left[V_1(b_{0|0}, B_{1|1}) | B_{0|0} = b_{0|0}, \tilde{E}_1 = \tilde{e}_1^*[b_{0|0}] \right] \quad (117)$$

Next, from (33) we have that $I(W_i; Y_i | Y^{i-1}) = \mathbb{E} [D(B_{i|i} | \Phi(B_{i-1|i-1}))]$ and thus from (V.1) we have:

$$J_{n, \pi^*}^\alpha = \mathbb{E} [V_0(S_0)] = \min_{e \in \mathcal{E}} -I(W^n; Y^n) + \alpha \mathbb{E}_e \left[\sum_{i=1}^n \eta(X_i) \right]. \quad (118)$$

Lastly, it follows directly that a more ‘concise’ sufficient statistic exists for the encoder - namely that it does not need to maintain z_{i-1} to produce x_{i+1} because under any optimal scheme, $Z_{i-1} = B_{i-1|i-1}$ and thus $\sigma(Z_{i-1}) \subset \sigma(B_{i|i})$ so the state variable $S_i = (Z_{i-1}, B_{i|i})$ can be reduced to $S_i = (B_{i|i})$ with Lemma IV.2 still holding. ■

APPENDIX F PROOF OF LEMMA VI.2

Proof: Note the following standard set of inequalities:

$$R_n(\rho, P_{W^n}, D) \leq \frac{1}{n} I(W^n; Z^n) \quad (119a)$$

$$\leq \frac{1}{n} I(W^n; Y^n) \quad (119b)$$

$$= \frac{1}{n} \sum_{i=1}^n I(W_i; Y_i | Y^{i-1}) \quad (119c)$$

$$= \frac{1}{n} \sum_{i=1}^n D(P_{Y_i|W_i, Y^{i-1}} \| P_{Y_i|Y^{i-1}} | P_{W_i, Y^{i-1}}) \quad (119d)$$

$$= \frac{1}{n} \sum_{i=1}^n D(P_{Y_i|X_i} \| P_{Y_i|Y^{i-1}} | P_{X_i, Y^{i-1}}) \quad (119e)$$

$$\leq \frac{1}{n} \sum_{i=1}^n D(P_{Y_i|X_i} \| P_{Y_i} | P_{X_i}) \quad (119f)$$

$$= \frac{1}{n} \sum_{i=1}^n I(X_i; Y_i) \quad (119g)$$

$$\leq \frac{1}{n} \sum_{i=1}^n C(\eta, P_{Y|X}, \mathbb{E}[\eta(X_i)]) \quad (119f)$$

$$\leq C(\eta, P_{Y|X}, L) \quad (119g)$$

where (119a) follows (43); (119b) follows from the data processing inequality; (119c) follows from Lemma V.1; (119d) follows from the definition of conditional mutual information (9) and the fact that X_i is a function of W_i and Y^{i-1} under policy \bar{e} ; (119e) follows from the memoryless nature of the channel (17) and Jensen’s inequality; (119f) follows from (12); and (119g) follows from (42) and the concavity of the capacity-cost function [59]. ■

APPENDIX G PROOF OF LEMMA VI.4

Proof: To prove (45a),

$$\begin{aligned} & P_{Z_i | Z^{i-1}=z^{i-1}, W^n=w^n}(dz_i) \\ &= \int_Y P_{Z_i | Z^{i-1}=z^{i-1}, W^n=w^n, Y_i=y}(dz_i) \\ & \quad \times P_{Y_i | Z^{i-1}=z^{i-1}, W^n=w^n}(dy) \\ &= \int_Y P_{Z_i | Z^{i-1}=z^{i-1}, W^n=w^n, Y_i=y}(dz_i) \\ & \quad \times P_{Y_i | Z^{i-1}=z^{i-1}, W^n=w^n, X_i=\bar{e}(w_i, z_{i-1})}(dy) \quad (120) \end{aligned}$$

$$= \int_Y 1_{\{z_i=\bar{d}(z_{i-1}, y_i)\}} P_{Y|X=\bar{e}(w_i, z_{i-1})}(dy) \quad (121)$$

$$\triangleq Q_{Z'|Z, W'}(dz_i | z_{i-1}, w_i) \quad (122)$$

where (120) follows from the stationary Markov encoder policy: $x_i = \bar{e}(w_i, z_{i-1})$; (121) follows from defining $1_{\{z_i=\bar{d}(z_{i-1}, y_i)\}}$ as a Dirac measure at the point $\bar{d}(z_{i-1}, y_i)$, the stationary Markov decoder policy $z_i = \bar{d}(z_{i-1}, y)$, and the non-anticipative and memoryless nature of the channel (17); and (122) simply denotes the time-invariant nature of the conditional distribution;

To prove (45b), we exploit the assumption that $\{Y_i\}$ are i.i.d. Because of this, we can denote $(Z_i : i = 1, \dots, n)$ by the following composition of independent random maps:

$$Z_i = \bar{d}(Z_{i-1}, Y_i) \triangleq \bar{d}_{Y_i}(Z_{i-1}) = \bar{d}_{Y_i} \circ \bar{d}_{Y_{i-1}} \circ \dots \circ \bar{d}_{Y_2}(Z_1)$$

This is thus an an iterated function system (IFS) [62], which is a time-homogeneous Markov chain over the state space \mathcal{Z} . ■

APPENDIX H PROOF OF LEMMA VII.2

Proof: Let $E_i \triangleq W_i - \mathbb{E}[W_i | Y^{i-1}]$ be the error term in estimation. We now select the statistics of W_0 such that $X_i \sim \mathcal{N}(0, L), \forall i$. The normalizing coefficient can be expressed as $\beta_i = \sqrt{\frac{L}{\text{Cov}(E_i, E_i)}}$, where the covariance of the error term can be recursively computed using

$$\text{Cov}(E_i, E_i) = \begin{cases} \frac{\rho^2 \sigma_n^2}{L + \sigma_n^2} \text{Cov}(E_{i-1}, E_{i-1}) + \sigma_m^2, & i \geq 1; \\ \text{Cov}(W_0, W_0), & i = 0 \end{cases} \quad (123)$$

Let the steady state value of the covariance from (123) be denoted by C . Then,

$$C \triangleq \frac{\sigma_m^2}{1 - \rho^2 \frac{\sigma_n^2}{L + \sigma_n^2}} \quad (124)$$

Note that because of the choice of \tilde{W}_0 in (76b), $\text{Cov}(E_i, E_i) = C$ and $\beta_i = \beta = \sqrt{\frac{L}{C}}$ for all $i \geq 0$.

Since all operations are linear and all primitive random variables $(\tilde{W}_i, V_i : i \geq 1)$ are i.i.d. and Gaussian, and since all other relationships are linear, all random variables are jointly Gaussian. From standard MMSE estimation theory, E_i thus independent of Y^{i-1} . As such, clearly $I(X_i; Y^{i-1}) = 0$. Since the initial condition \tilde{W}_0 is chosen according to (76a), $X_i \sim \mathcal{N}(0, L)$ for all i . Therefore, since the variance of V_i 's is σ_v^2 , this means that Y 's are i.i.d. The policies (78) are thus stationary-Markov coordination strategies:

$$X_i = \beta (W_i - \rho Z_{i-1}) \quad (125a)$$

$$Z_i = \rho Z_{i-1} + \gamma Y_i \quad (125b)$$

where (125a) follows because $\mathbb{E}[W_i | Y^{i-1}] = \mathbb{E}[\rho W_{i-1} + \tilde{W}_i | Y^{i-1}] = \rho Z_{i-1}$, and (125b) follows by expanding $\mathbb{E}[W | Y^i]$ using the innovation sequence and exploiting how Y_i are i.i.d. The value of the parameters β, γ are given by

$$\beta = \sqrt{\frac{L}{C}}, \gamma = \frac{\beta C}{L + \sigma_n^2} \quad (126)$$

Note that from the definition of C in (124), $P_{W_i | Z_{i-1}=z_{i-1}} \sim \mathcal{N}(\rho z_{i-1}, C)$. Hence, using (79),

$$\begin{aligned} Q_{Z' | Z, W'}(\cdot | z_{i-1}, w_i) &\sim \mathcal{N}(\rho z_{i-1} + \beta \gamma (w_i - \rho z_{i-1}), \gamma^2 \sigma_v^2) \\ Q_{Z' | Z}(\cdot | z_{i-1}) &\sim \mathcal{N}(\rho z_{i-1}, \gamma^2 (L + \sigma_v^2)) \end{aligned}$$

From Theorem VI.5, the linear stationary Markov coordination strategy (78) is inverse control optimal for a ρ of the form

$$\begin{aligned} \rho(w_i, z_{i-1}, z_i) &\propto_+ -\log \frac{dQ_{Z' | Z, W'}(\cdot | z_{i-1}, w_i)}{dQ_{Z' | Z}(\cdot | z_{i-1})}(z_i) \\ &= \frac{(z_i - \rho z_{i-1} - \beta \gamma (w_i - \rho z_{i-1}))^2}{2\gamma^2 \sigma_v^2} \\ &\quad - \frac{(z_i - \rho z_{i-1})^2}{2\gamma^2 (L + \sigma_v^2)} - \log \sqrt{\frac{L + \sigma_v^2}{\sigma_v^2}} \\ &\propto_+ (z_i - w_i)^2 - \frac{\sigma_v^2}{L + \sigma_v^2} (w_i - \rho z_{i-1})^2 \end{aligned} \quad (127)$$

where (127) follows from (126). Similarly, the power-like cost for inverse control optimality is given $\eta(x) \propto_+ D(P_{Y|X=x} \| P_Y) = D(P_V(\cdot - x) \| P_Y(\cdot)) \propto_+ x^2$. Thus we

have

$$\begin{aligned} &\mathbb{E} \left[\sum_{i=1}^n \rho (W_i, Z_{i-1}, Z_i) \right] \\ \propto_+ &\mathbb{E} \left[\sum_{i=1}^n (W_i - Z_i)^2 - \left(\frac{\sigma_v^2}{L + \sigma_v^2} \right) (W_i - \rho Z_{i-1})^2 \right] \end{aligned} \quad (128)$$

$$\begin{aligned} &= \mathbb{E} \left[\sum_{i=1}^n (W_i - Z_i)^2 \right] \\ &\quad - \mathbb{E} \left[\left(\frac{\sigma_v^2}{L + \sigma_v^2} \right) (\rho W_{i-1} - \rho Z_{i-1} + \tilde{W}_i)^2 \right] \end{aligned} \quad (129)$$

$$\begin{aligned} &= \mathbb{E} \left[\sum_{i=1}^n (W_i - Z_i)^2 - \left(\frac{\sigma_v^2 \rho^2}{L + \sigma_v^2} \right) (W_{i-1} - \rho Z_{i-1})^2 \right] \\ &\quad - \mathbb{E} \left[\left(\frac{\sigma_v^2}{L + \sigma_v^2} \right) \tilde{W}_i^2 \right] \end{aligned} \quad (130)$$

$$\begin{aligned} &= \mathbb{E} \left[\sum_{i=1}^n \left(1 - \frac{\sigma_v^2 \rho^2}{L + \sigma_v^2} \right) (W_i - Z_i)^2 - \left(\frac{\sigma_v^2}{L + \sigma_v^2} \right) \tilde{W}_i^2 \right] \\ &\quad - \frac{\sigma_v^2 \rho^2}{L + \sigma_v^2} \mathbb{E}[W_0^2] + \frac{\sigma_v^2 \rho^2}{L + \sigma_v^2} \mathbb{E}[(Z_n - W_n)^2] \\ \propto_+ &\mathbb{E} \left[\sum_{i=1}^n (W_i - Z_i)^2 \right] + \left(\frac{1}{1 - \frac{\sigma_v^2 \rho^2}{L + \sigma_v^2}} \right) \mathbb{E}[(Z_n - W_n)^2] \end{aligned}$$

where (128) follows from (80b); (129) follows from (76b); (130) follows from (76c). ■

REFERENCES

- [1] C. Papadimitriou and J. Tsitsiklis, "Intractable problems in control theory," *SIAM Journal on Control and Optimization*, vol. 24, no. 4, pp. 639–654, 1986.
- [2] T. M. Cover and J. Thomas, *Elements of Information Theory*, 2nd ed. New York: Wiley, 2006.
- [3] D. Bertsekas and S. Shreve, *Stochastic optimal control: The discrete time case*. Athena Scientific, 1978.
- [4] M. Gastpar, B. Rimoldi, and M. Vetterli, "To code, or not to code: Lossy source-channel communication revisited," *IEEE Transactions on Information Theory*, vol. 49, no. 5, pp. 1147–1158, May 2003.
- [5] R. Kalman, "When is a linear control system optimal?" *Trans. ASME Ser. D:J: Basic Eng*, pp. 1–10, 1964.
- [6] K. Kording and D. Wolpert, "The loss function of sensorimotor learning," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 101, no. 26, p. 9839, 2004.
- [7] E. Todorov, "Optimality principles in sensorimotor control," *Nature neuroscience*, vol. 7, no. 9, pp. 907–915, 2004.
- [8] C. Baker, J. Tenenbaum, and R. Saxe, "Goal inference as inverse planning," in *Proceedings of the 29th annual meeting of the cognitive science society*, 2007.
- [9] C. Shannon, "A mathematical theory of Communication," *Bell System Technical Journal*, vol. 22, pp. 379–423, 1948.
- [10] H. Witsenhausen, "Separation of estimation and control for discrete time systems," *Proceedings of the IEEE*, vol. 59, pp. 1557–1566, 1971.
- [11] —, "Informational aspects of stochastic control," in *Proc. the Oxford Conference on Stochastic Optimization*, 1978.
- [12] D. Teneketzis, "Communication in decentralized control," Ph.D. dissertation, Massachusetts Institute of Technology, 1980.
- [13] O. Shayevitz and M. Feder, "Communication with feedback via posterior matching," in *IEEE International Symposium on Information Theory*, Nice, France, June 2007.
- [14] —, "The posterior matching feedback scheme: Capacity achieving and error analysis," in *IEEE International Symposium on Information Theory*, Toronto, Canada, July 2008.
- [15] —, "Optimal Feedback Communication via Posterior Matching," *IEEE Transactions on Information Theory*, 2009, to appear, Arxiv preprint arXiv:0909.4828.

- [16] M. Horstein, "Sequential transmission using noiseless feedback," *IEEE Transactions on Information Theory*, vol. 9, no. 3, pp. 136–143, 1963.
- [17] J. Schalkwijk and T. Kailath, "A coding scheme for additive noise channels with feedback—I: No bandwidth constraint," *IEEE Transactions on Information Theory*, vol. 12, no. 2, pp. 172–182, 1966.
- [18] J. Schalkwijk, "A coding scheme for additive noise channels with feedback—II: Band-limited signals," *IEEE Transactions on Information Theory*, vol. 12, no. 2, pp. 183–189, 1966.
- [19] T. P. Coleman, "A Stochastic Control Approach to 'Posterior Matching'-style Feedback Communication Schemes," in *IEEE International Symposium on Information Theory*, Seoul, Korea, July 2009.
- [20] S. Tatikonda, "Control Under Communication Constraints," Ph.D. dissertation, Massachusetts Institute of Technology, 2000.
- [21] S. Tatikonda, A. Sahai, and S. Mitter, "Stochastic linear control over a communication channel," *Automatic Control, IEEE Transactions on*, vol. 49, no. 9, pp. 1549–1561, 2004.
- [22] S. Tatikonda and S. Mitter, "Control under communication constraints," *Automatic Control, IEEE Transactions on*, vol. 49, no. 7, pp. 1056–1068, 2004.
- [23] N. Elia, "When Bode meets Shannon: Control-oriented feedback communication schemes," *IEEE transactions on Automatic control*, vol. 49, no. 9, pp. 1477–1488, 2004.
- [24] N. Martins and M. Dahleh, "Feedback control in the presence of noisy channels: Bode-like fundamental limitations of performance," *Automatic Control, IEEE Transactions on*, vol. 53, no. 7, pp. 1604–1615, 2008.
- [25] J. Liu and N. Elia, "Convergence of fundamental limitations in feedback communication, estimation, and feedback control over Gaussian channels," *Arxiv preprint arXiv:0910.0320*, 2009.
- [26] E. Ardestanizadeh and M. Franceschetti, "Control-theoretic Approach to Communication with Feedback: Fundamental Limits and Code Design," *Arxiv preprint arXiv:1006.5265*, 2010.
- [27] O. Shayevitz, "Posterior matching variants and fixed-point elimination," in *47th Annual Allerton Conference on Communication, Control, and Computing*. IEEE, 2009, pp. 935–939.
- [28] N. Merhav and M. Feder, "Universal prediction," *Information Theory, IEEE Transactions on*, vol. 44, no. 6, pp. 2124–2147, 2002.
- [29] N. Cesa-Bianchi and G. Lugosi, *Prediction, learning, and games*. Cambridge Univ Pr, 2006.
- [30] A. Aiyer, K. Pyun, Y. Huang, D. O'Brien, and R. Gray, "Lloyd clustering of Gauss mixture models for image compression and classification," *Signal Processing: Image Communication*, vol. 20, no. 5, pp. 459–485, 2005.
- [31] S. Mitter and N. Newton, "A variational approach to nonlinear estimation," *SIAM Journal on Control and Optimization*, vol. 42, no. 5, pp. 1813–1833, 2004.
- [32] —, "Information and entropy flow in the Kalman–Bucy filter," *Journal of Statistical Physics*, vol. 118, no. 1, pp. 145–176, 2005.
- [33] E. Todorov, "Efficient computation of optimal actions," *Proceedings of the National Academy of Sciences*, vol. 106, no. 28, p. 11478, 2009.
- [34] S. Tatikonda and S. Mitter, "The Capacity of Channels With Feedback," *IEEE Transactions on Information Theory*, vol. 55, no. 1, pp. 323–349, 2009.
- [35] J. Bae and A. Anastasopoulos, "A posterior matching scheme for finite-state channels with feedback," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*. IEEE, 2010, pp. 2338–2342.
- [36] M. Anand and P. R. Kumar, "Estimating the state of a Markov chain over a noisy communication channel: a bound and an encoder," in *IEEE Conference on Decision and Control*, Atlanta, GA, December 2010, to appear.
- [37] J. C. Walrand and P. Varaiya, "Optimal Causal Coding-Decoding Problems," *Information Theory, IEEE Transactions on*, vol. 29, no. 6, pp. 814–820, 1983.
- [38] A. Mahajan and D. Teneketzis, "On the design of globally optimal communication strategies for real-time noisy communication systems with noisy feedback," *Selected Areas in Communications, IEEE Journal on*, vol. 26, no. 4, pp. 580–595, 2008.
- [39] R. Bansal and T. Basar, "Simultaneous design of communication and control strategies for stochastic systems with feedback," *Automatica*, vol. 45, no. 5, pp. 679–694, 1989.
- [40] V. Borkar, S. Mitter, and S. Tatikonda, "Optimal sequential vector quantization of Markov sources," *SIAM journal on control and optimization*, vol. 40, no. 1, pp. 135–148, 2001.
- [41] S. Yuksel and T. Basar, "Causal coding of markov sources with continuous alphabets," *Analysis and Design of Nonlinear Control Systems*, pp. 91–108, 2008.
- [42] J. Casti, "On the general inverse problem of optimal control theory," *Journal of Optimization Theory and Applications*, vol. 32, no. 4, pp. 491–497, 1980.
- [43] A. Ng and S. Russell, "Algorithms for inverse reinforcement learning," in *Proceedings of the Seventeenth International Conference on Machine Learning*. Morgan Kaufmann Publishers Inc., 2000, pp. 663–670.
- [44] P. Abbeel and A. Ng, "Apprenticeship learning via inverse reinforcement learning," in *Proceedings of the twenty-first international conference on Machine learning*. ACM, 2004, p. 1.
- [45] K. Dvijotham and E. Todorov, "Inverse Optimal Control with Linearly-Solvable MDPs," in *Proceedings of the International Conference on Machine Learning*, 2010.
- [46] P. Abbeel, A. Coates, M. Quigley, and A. Ng, "An application of reinforcement learning to aerobatic helicopter flight," in *Advances in Neural Information Processing Systems 19: Proceedings of the 2006 Conference*. The MIT Press, 2007.
- [47] I. Csiszár, "On an extremum problem of information theory," *Studia Scientiarum Mathematicarum Hungarica*, vol. 9, no. 1, pp. 57–71, 1974.
- [48] I. Csiszar and J. Korner, *Information theory: coding theorems for discrete memoryless systems*. Academic Press, Inc. Orlando, FL, USA, 1982.
- [49] M. Gastpar, "To Code or Not to Code," Ph.D. dissertation, EPFL, 2002.
- [50] P. Chigansky, R. Liptser, and R. Van Handel, "Intrinsic methods in filter stability," *Handbook of Nonlinear Filtering*, 2009.
- [51] M. Propp, "The thermodynamic properties of Markov processes," Ph.D. dissertation, M.I.T, 1985.
- [52] R. Elliott, L. Aggoun, and J. Moore, *Hidden Markov models: estimation and control*. Springer, 1995.
- [53] C. Omar, A. Akce, M. Johnson, T. Bretl, R. Ma, E. Maclin, M. McCormick, and T. P. Coleman, "A Feedback Information-Theoretic Approach to the Design of Brain-Computer Interfaces," *Int'l Journal on Human-Computer Interaction*, January 2011, special issue on Current Trends in Brain-Computer Interface (BCI) Research.
- [54] V. Anantharam and S. Verdú, "Bits through queues," *IEEE Transactions on Information Theory*, vol. 42, no. 1, pp. 4–18, 1996.
- [55] R. Sundaresan and S. Verdú, "Capacity of queues via point-process channels," vol. 52, no. 6, pp. 2697–2709, June 2006.
- [56] D. Blackwell, *Information Theory*, in *Modern Mathematics for the Engineer*, 2nd ed. McGraw-Hill, 1961.
- [57] R. Ahlswede and A. Kaspi, "Optimal coding strategies for certain permuting channels," *IEEE Trans. on Information Theory*, vol. 33, no. 3, pp. 310–314, 1987.
- [58] H. Permuter, P. Cuff, B. Roy, and T. Weissman, "Capacity of the Trapdoor Channel with Feedback," *IEEE Trans. on Information Theory*, pp. 3150–65, July, 2008.
- [59] R. McEliece, *The theory of information and coding*. Cambridge Univ Press, 2002.
- [60] A. Mahajan, "Sequential decomposition of sequential dynamic teams: applications to real-time communication and networked control systems," Ph.D. dissertation, University of Michigan, 2008.
- [61] D. Bertsekas, *Nonlinear programming*. Athena Scientific Belmont, MA, 1999.
- [62] Y. Kifer, *Ergodic Theory of Random Transformations*. Boston: Birkhäuser, 1986.
- [63] F. P. Kelly, *Reversibility and Stochastic Networks*. New York: Kluwer Academic Publishers, 1979.
- [64] R. G. Gallager, *Discrete Stochastic Processes*, 1st ed. Boston: Kluwer Academic Publishers, 1996.
- [65] S. Meyn and R. Tweedie, *Markov Chains and Stochastic Stability*, 2nd ed. New York: Cambridge University Press, 2009.
- [66] J. Alvarez and B. Hajek, "Equivalence of trans paths in ion channels," *Physical Review E*, vol. 73, no. 4, 2006.
- [67] J. Hsu and P. Burke, "Behavior of tandem buffers with geometric input and Markovian output," *IEEE Transactions on Communications*, vol. 24, no. 3, pp. 358–361, 1976.
- [68] T. P. Coleman, "A simple memoryless proof of the capacity of the exponential server timing channel," in *Proceedings of the IEEE Information Theory Workshop (ITW)*, Volos, Greece, 2009.
- [69] J. Martin, "Batch queues, reversibility and first-passage percolation," *Queueing Systems*, vol. 62, no. 4, pp. 411–427, 2009.
- [70] M. S. McCormick, R. Ma, and T. P. Coleman, "An Analytic Spatial Filter and Hidden Markov Model for Enhanced Information Transfer Rate in EEG-Based Brain Computer Interfaces," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Dallas, TX, 2010.

- [71] S. Wills and D. MacKay, "DASHER-an efficient writing system for brain-computer interfaces?" *Neural Systems and Rehabilitation Engineering, IEEE Transactions on*, vol. 14, no. 2, pp. 244–246, 2006.
- [72] A. Akce, M. Johnson, and T. Bretl, "Remote Teleoperation of an Unmanned Aircraft with a Brain-Machine Interface: Theory and Preliminary Results," in *Robot. and Autom. Proceedings. IEEE Int. Conf. on*, 2010.
- [73] S. Gorantla and T. Coleman, "On reversible Markov chains and maximization of directed information," in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*. IEEE, 2010, pp. 216–220.
- [74] K. Friston, "The free-energy principle: a unified brain theory?" *Nature Reviews Neuroscience*, vol. 11, no. 2, pp. 127–138, 2010.
- [75] S. Mitter, "Towards a unified view of inference, communication and control," in *Signal Processing and Communications (SPCOM), 2010 International Conference on*. IEEE, 2010, pp. 1–2.
- [76] T. Lee and D. Mumford, "Hierarchical Bayesian inference in the visual cortex," *JOSA A*, vol. 20, no. 7, pp. 1434–1448, 2003.
- [77] R. Rao and D. Ballard, "Predictive coding in the visual cortex: a functional interpretation of some extra-classical receptive-field effects," *Nature Neuroscience*, vol. 2, pp. 79–87, 1999.
- [78] I. Fiete, Y. Burak, and T. Brookings, "What grid cells convey about rat location," *Journal of Neuroscience*, vol. 28, no. 27, p. 6858, 2008.
- [79] Y. Burak and I. Fiete, "Accurate path integration in continuous attractor network models of grid cells," *PLoS computational biology*, vol. 5, no. 2, pp. 801–806, 2009.
- [80] R. Dudley, *Real analysis and probability*. Cambridge Univ Pr, 2002.